

HANDBOOK OF DIFFERENTIAL EQUATIONS

Evolutionary Equations
VOLUME 3

Edited by
C.M. Dafermos
E. Feireisl



VISIT...

LANZAROTE
Caliente.COM

HANDBOOK OF DIFFERENTIAL EQUATIONS: EVOLUTIONARY EQUATIONS, 3

Edited by

C. Dafermos, Brown University, Providence, USA

Eduard Feireisl, Mathematical Institute AS CR, Prague, Czech Republic.

Description

The material collected in this volume reflects the active present of this area of mathematics, ranging from the abstract theory of gradient flows to stochastic representations of non-linear parabolic PDE's. Articles will highlight the present as well as expected future directions of development of the field with particular emphasis on applications. The article by Ambrosio and Savare discusses the most recent development in the theory of gradient flow of probability measures. After an introduction reviewing the properties of the Wasserstein space and corresponding subdifferential calculus, applications are given to evolutionary partial differential equations. The contribution of Herrero provides a description of some mathematical approaches developed to account for quantitative as well as qualitative aspects of chemotaxis. Particular attention is paid to the limits of cell's capability to measure external cues on the one hand, and to provide an overall description of aggregation models for the slim mold *Dictyostelium discoideum* on the other. The chapter written by Masmoudi deals with a rather different topic - examples of singular limits in hydrodynamics. This is nowadays a well-studied issue given the amount of new results based on the development of the existence theory for rather general systems of equations in hydrodynamics. The paper by DeLellis addresses the most recent results for the transport equations with regard to possible applications in the theory of hyperbolic systems of conservation laws. Emphasis is put on the development of the theory in the case when the governing field is only a BV function. The chapter by Rein represents a comprehensive survey of results on the Poisson-Vlasov system in astrophysics. The question of global stability of steady states is addressed in detail. The contribution of Soner is devoted to different representations of non-linear parabolic equations in terms of Markov processes. After a brief introduction on the linear theory, a class of non-linear equations is investigated, with applications to stochastic control and differential games. The chapter written by Zuazua presents some of the recent progresses done on the problem of controllability of partial differential equations. The applications include the linear wave and heat equations, parabolic equations with coefficients of low regularity, and some fluid-structure interaction models.

Contents

Preface

Contributors	vii
1.L. Ambrosio, G. Savare: Gradient flows of probability measures	1
2.M.A. Herrero: The mathematics of chemotaxis	137
3.N. Masmoudi: Examples of singular limits in hydrodynamics	195
4. C. DeLellis: Notes on hyperbolic systems of conservation laws and transport equations	277
5. G. Rein: Collisionless kinetic equations from astrophysics - the Vlasov-Poisson system	383
6. H.M. Soner: Stochastic representations for non-linear parabolic PDE's	477
7. E. Zuazua: Controllability and observability of partial differential equations: Some results and open problems	527
Index	623

Hardbound, 652 pages, publication date: OCT-2006
ISBN-13: 978-0-444-52848-3
ISBN-10: 0-444-52848-2

Preface

The original aim of this series of *Handbook of Differential Equations* was to acquaint the interested reader with the current status of the theory of evolutionary partial differential equations, with regard to some of its applications in physics, biology, chemistry, economy, among others. The material collected in this volume reflects the active present of this area of mathematics, ranging from the abstract theory of gradient flows to stochastic representations of nonlinear parabolic PDEs.

The aim here is to collect review articles, written by leading experts, which will highlight the present as well as expected future directions of development of the field with particular emphasis on applications. The contributions are presented in alphabetical order according to the name of the first author. The article by Ambrosio and Savaré discusses the most recent development in the theory of gradient flow of probability measures. After an introduction reviewing the properties of the Wasserstein space and corresponding subdifferential calculus, applications are given to evolutionary partial differential equations. The contribution of Herrero provides a description of some mathematical approaches developed to account for quantitative as well as qualitative aspects of chemotaxis. Particular attention is paid to the limits of cell's capability to measure external cues on the one hand, and to provide an overall description of aggregation models for the slim mold *Dictyostelium discoideum* on the other. The chapter written by Masmoudi deals with a rather different topic – examples of singular limits in hydrodynamics. This is nowadays a well-studied issue given the amount of new results based on the development of the existence theory for rather general systems of equations in hydrodynamics. The chapter by De Lellis addresses the most recent results for the transport equations with regard to possible applications in the theory of hyperbolic systems of conservation laws. Emphasis is put on the development of the theory in the case when the governing field is only a BV function. The chapter by Rein represents a comprehensive survey of results on the Poisson–Vlasov system in astrophysics. The question of global stability of steady states is addressed in detail. The contribution of Soner is devoted to different representations of nonlinear parabolic equations in terms of Markov processes. After a brief introduction on the linear theory, a class of nonlinear equations is investigated, with applications to stochastic control and differential games. The chapter written by Zuazua presents some of the recent progresses done on the problem of controllability of partial differential equations. The applications include the linear wave and heat equations, parabolic equations with coefficients of low regularity, and some fluid–structure interaction models.

We firmly believe that the fascinating variety of rather different topics covered by this volume will contribute to inspiring and motivating researchers in the future.

Constantine Dafermos
Eduard Feireisl

List of Contributors

- Ambrosio, L., *Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, 56126 Pisa, Italy*
(Ch. 1)
- De Lellis, C., *Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland* (Ch. 4)
- Herrero, M.A., *Departamento de Matemática Aplicada, Facultad de CC. Matemáticas, Universidad Complutense de Madrid, Avda. Complutense s/n, 28040 Madrid, Spain*
(Ch. 2)
- Masmoudi, N., *Courant Institute, New York University, 251 Mercer Street, New York, NY 10012-1185, USA* (Ch. 3)
- Rein, G., *Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany*
(Ch. 5)
- Savaré G., *Dipartimento di Matematica, Università di Pavia, Pavia via Ferrata 1, 27100 Pavia, Italy* (Ch. 1)
- Soner, H.M., *Koç University, Istanbul, Turkey* (Ch. 6)
- Zuazua, E., *Departamento de Matemáticas, Universidad Autónoma, 28049 Madrid, Spain*
(Ch. 7)

CHAPTER 1

Gradient Flows of Probability Measures

Luigi Ambrosio

Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, 56126 Pisa, Italy

E-mail: l.ambrosio@sns.it

Giuseppe Savaré

Dipartimento di Matematica, Università di Pavia, Pavia via Ferrata 1, 27100 Pavia, Italy

E-mail: giuseppe.savare@unipv.it

Contents

Introduction	3
Notation	7
1. Notation and measure-theoretic results	8
1.1. Transport maps and transport plans	9
1.2. Narrow convergence	10
1.3. The change of variables formula	11
2. Metric and differentiable structure of the Wasserstein space	13
2.1. Absolutely continuous maps and metric derivative	13
2.2. The quadratic optimal transport problem	14
2.3. Geodesics in $\mathcal{P}_2(\mathbb{R}^d)$	16
2.4. Existence of optimal transport maps	17
2.5. The continuity equation with locally Lipschitz velocity fields	19
2.6. The tangent bundle to the Wasserstein space	29
3. Convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$	38
3.1. λ -geodesically convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$	39
3.2. Examples of convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$	40
3.3. Relative entropy and convex functionals of measures	47
3.4. Log-concavity and displacement convexity	50
4. Subdifferential calculus in $\mathcal{P}_2(\mathbb{R}^d)$	55
4.1. Definition of the subdifferential for a.c. measures	58
4.2. Subdifferential calculus in $\mathcal{P}_2^a(\mathbb{R}^d)$	60
4.3. The case of λ -convex functionals along geodesics	62
4.4. Regular functionals	65
4.5. Examples of subdifferentials	68

HANDBOOK OF DIFFERENTIAL EQUATIONS

Evolutionary Equations, volume 3

Edited by C.M. Dafermos and E. Feireisl

© 2007 Elsevier B.V. All rights reserved

5. Gradient flows of λ -geodesically convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$	84
5.1. Characterizations of gradient flows, uniqueness and contractivity	85
5.2. Main properties of gradient flows	89
5.3. Existence of gradient flows by convergence of the “minimizing movement” scheme	95
5.4. Bibliographical notes	104
6. Applications to evolution PDEs	107
6.1. Gradient flows and evolutionary PDEs of diffusion type	107
6.2. The linear transport equation for λ -convex potentials	111
6.3. Kolmogorov–Fokker–Planck equation	113
6.4. Nonlinear diffusion equations	129
6.5. Drift diffusion equations with nonlocal terms	132
6.6. Gradient flow of $-W^2/2$ and geodesics	133
References	133

Introduction

In a finite-dimensional smooth setting, the gradient flow of a function $\phi : \mathbb{M}^d \rightarrow \mathbb{R}$ defined on a Riemannian manifold $\mathbb{M}^d$ simply means the family of solutions $u : \mathbb{R} \rightarrow \mathbb{M}^d$ of the Cauchy problem associated to the differential equation

$$\frac{d}{dt}u(t) = -\nabla\phi(u(t)) \quad \text{in } T_{u(t)}\mathbb{M}^d, t \in \mathbb{R}; \quad u(0) = u_0 \in \mathbb{M}^d. \quad (0.1)$$

Thus, at each time $t \in \mathbb{R}$ equation (0.1), which is imposed in the tangent space $T_{u(t)}\mathbb{M}^d$ of $\mathbb{M}^d$ at the moving point $u(t)$, simply prescribes that the velocity vector $\mathbf{v}_t := \frac{d}{dt}u(t)$ of the curve u equals the opposite of the gradient of ϕ at $u(t)$.

The extension of the theory of gradient flows to suitable (infinite-dimensional) abstract/functional spaces and its link with evolutionary PDEs is a wide subject with a long history.

One of its first main achievement, going back to the pioneering papers by Komura [61], Crandall and Pazy [33], Brézis [21] (we refer to the monograph [22]), concerns an Hilbert space H and nonlinear contraction semigroups generated by a proper, convex, and lower semicontinuous functional $\phi : H \rightarrow (-\infty, +\infty]$. Since in general ϕ admits only a subdifferential $\partial\phi$ in a (possibly strict) subset $D(\partial\phi) \subset D(\phi) := \{u \in H : \phi(u) < +\infty\}$ and each tangent space of H can be identified with H itself, it turns out that (0.1) should be rephrased as a *subdifferential inclusion* on the positive real line

$$u'(t) \in -\partial\phi(u(t)), \quad t > 0; \quad u(0) = u_0 \in \overline{D(\phi)}, \quad (0.2)$$

and it provides a general framework for studying existence, uniqueness, stability, asymptotic behavior, and regularizing properties of many PDEs of parabolic type.

The possibility to work in a more general metric space (E, d) and/or with non-smooth perturbations of a convex functional $\phi : E \rightarrow (-\infty, +\infty]$ has been exploited by De Giorgi and his collaborators in a series of papers originating from [37] and culminating in [64] (see also the presentation of [6] and our recent book [9]). One of the nice features of this approach is the so-called “minimizing movement” approximation scheme [36]: it suggests a general variational procedure to approximate and construct gradient flows by a recursive minimization algorithm. For, one introduces a uniform partition $0 < \tau < 2\tau < \dots < n\tau < \dots$ of the positive real line, $\tau > 0$ being the step size, and starting from the initial value $U_\tau^0 := u_0$ one looks for a suitable approximation U_τ^n of u at the time $n\tau$ by iteratively solving the minimum problems

$$\min_{U \in E} \phi(U) + \frac{1}{2\tau} d^2(U, U_\tau^{n-1}). \quad (0.3)$$

Under general lower semicontinuity and coercivity assumptions, a minimizer U_τ^n of (0.3) exists so that a piecewise constant interpolant U_τ taking the value U_τ^n in each interval $((n-1)\tau, n\tau]$ can be constructed. Limit points (possibly after extracting a suitable subsequence) of $U_\tau(t)$ as $\tau \downarrow 0$ can be considered as good candidates for gradient flows of ϕ and

in many circumstances it is, in fact, possible to give differential characterizations of their trajectories.

One of the most striking application of this variational point of view has been introduced by Otto [57,74] (also in collaboration with Jordan and Kinderlehrer): he showed that the Fokker–Planck equation

$$\partial_t u - \nabla \cdot (\nabla u + u \nabla V) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty) \quad (0.4)$$

and nonlinear diffusion equations of porous media type

$$\partial_t u - \Delta \beta(u) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty) \quad (0.5)$$

can be interpreted as gradient flows, in the metric space $E := \mathcal{P}_2(\mathbb{R}^d)$ of Borel probability measures in $\mathbb{R}^d$ with finite quadratic moment, of suitable integral functionals of the type

$$\phi(\mu) := \int_{\mathbb{R}^d} F(\rho(x)) d\gamma(x), \quad \rho := \frac{d\mu}{d\gamma}, \quad (0.6)$$

for a suitable choice of the nonlinearity F and of the reference measure γ in $\mathbb{R}^d$. Here the solutions u_t of (0.4) and (0.5) yield a corresponding family of evolving measures $\mu_t \in \mathcal{P}_2(\mathbb{R}^d)$ through the identification $\mu_t = u_t \mathcal{L}^d$.

One of the main novelties of Otto’s approach relies in the particular *distance* d on $\mathcal{P}_2(\mathbb{R}^d)$ which should be used to recover the above mentioned PDEs in the limit: it is the so-called Kantorovich–Rubinstein–Wasserstein distance between two measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, defined as

$$W_2^2(\mu, \nu) := \min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) : \right. \\ \left. \gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_{\#}^1 \gamma = \mu, \pi_{\#}^2 \gamma = \nu \right\}. \quad (0.7)$$

The minimum in (0.7) is thus evaluated on all probability measures γ on the product $\mathbb{R}^d \times \mathbb{R}^d$ whose marginals $\pi_{\#}^1 \gamma, \pi_{\#}^2 \gamma$ are μ and ν , respectively, $\pi^1, \pi^2: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ denote the canonical projections on the first and the second factor.

By applying the “minimizing movement” scheme in $\mathcal{P}_2(\mathbb{R}^d)$ with the above choice (0.6) of ϕ and with $d := W_2$, it is, in fact, possible to show that its discrete trajectories converge to the solution of a suitable evolution PDE. Moreover, Otto introduced a formal “Riemannian” structure in the space $\mathcal{P}_2(\mathbb{R}^d)$ in order to guess first, and then prove rigorously the form of the limit PDEs and their gradient flow structure like in (0.1).

The aim of this chapter is to present, in a simplified form, the general and rigorous theory developed in our book [9] (written with N. Gigli), giving quite general answers to the following questions:

1. Give a rigorous meaning to the concept of gradient flow in $\mathcal{P}_2(\mathbb{R}^d)$.
2. Find general conditions on ϕ in order to guarantee the convergence of the “minimizing movement” scheme in $\mathcal{P}_2(\mathbb{R}^d)$.

3. Characterize the limit trajectories and study their properties, applying them to classes of specific and relevant examples.

In comparison with [9], the simplification comes from the fact that we mostly restrict ourselves to absolutely continuous measures, in finite-dimensional spaces, while in [9] none of these restrictions is present.

Concerning the first point, it is clear from the heuristic arguments of Otto and from (0.1) that one should make precise:

- (1a) the notion of *velocity vector field* of a curve $(\mu_t)_{t \in (0, T)}$ of measures in $\mathcal{P}_2(\mathbb{R}^d)$,
- (1b) the notion of *tangent space* $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ of $\mathcal{P}_2(\mathbb{R}^d)$ at a given measure μ ,
- (1c) the notion of *gradient* of a functional ϕ (like (0.6)) at μ .

The investigations about velocity and tangent space are, in fact, strictly related to a deep analysis of the *continuity equation*

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \quad \text{in } \mathbb{R}^d \times (0, T).$$

It is carried out in Section 2.6 after some basic preliminaries of measure theory (recalled in Section 1), a brief outline on optimal transportation and Wasserstein distance (presented in Sections 2.1–2.4), and a more detailed review on the classical representation formulas for solutions of the continuity equation, which is discussed in Section 2.5. Starting from the general definition of absolutely continuous curves in a (arbitrary) metric space, we will show that every absolutely continuous family of measures $(\mu_t)_{t \in (0, T)}$ in $\mathcal{P}_2(\mathbb{R}^d)$ satisfies the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \quad \text{in the distribution sense of } \mathcal{D}'(\mathbb{R}^d \times (0, T)), \quad (0.8)$$

for a suitable *Borel velocity vector field* $\mathbf{v}_t \in L^2(\mu_t; \mathbb{R}^d)$ satisfying

$$\text{Length}_a^b(\mu_t) = \int_0^b \left(\int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 d\mu_t(x) \right)^{1/2} dt \quad \forall 0 \leq a < b \leq T. \quad (0.9)$$

Furthermore, (0.8) and (0.9) uniquely determine $\mathbf{v}_t$ in $L^2(\mu_t; \mathbb{R}^d)$ up to a negligible set of times.

Since $\mathcal{P}_2(\mathbb{R}^d)$ is a length space (i.e., the infimum of the distance between any two points is the infimum of the lengths of all curves connecting the two points), one recovers also the Benamou–Brenier [15] formula

$$W_2(\mu, \nu) = \min \left\{ \int_0^1 \left(\int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 d\mu_t(x) \right)^{1/2} dt : \right. \\ \left. \mu_t \in AC((0, 1); \mathcal{P}_2(\mathbb{R}^d)) \text{ satisfies (0.8), } \mu_0 = \mu, \mu_1 = \nu \right\}. \quad (0.10)$$

Recalling the usual definition of the Riemannian distance on a manifold, we can thus consider $\mathbf{v}_t$ as the *velocity vector* of the curve (μ_t) and the squared $L^2(\mu_t; \mathbb{R}^d)$ -norm as the metric tensor in $\mathcal{P}_2(\mathbb{R}^d)$.

It turns out that in general the set spanned by all the possible velocity vector fields of a curve through a measure μ is a *proper* subset of $L^2(\mu; \mathbb{R}^d)$. For, $\mathbf{v}_t$ can be strongly approximated in $L^2(\mu_t; \mathbb{R}^d)$ by gradients of smooth functions (and this approximability property is equivalent to (0.9)); moreover, gradients of smooth functions are always velocity vectors (in the above sense) of smooth curves. These facts suggests the definition of the *tangent space* as

$$\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) := \overline{\{\nabla \varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu_t; \mathbb{R}^d)}. \quad (0.11)$$

One of the important byproducts of this analysis is the formula

$$\frac{d}{dt} W_2^2(\mu_t, \nu) = 2 \int_{\mathbb{R}^d} \langle \mathbf{v}_t, \mathbf{t}_{\mu_t}^\nu - \mathbf{i} \rangle d\mu_t \quad \text{for a.e. } t, \quad (0.12)$$

for the squared Wasserstein distance from a given measure ν . Here $\mathbf{t}_{\mu_t}^\nu$ are the optimal transport maps between μ_t and ν (provided they exist, as it happens whenever μ_t are absolutely continuous) and $\mathbf{i}$ is the identity map.

Concerning (1c), any reasonable definition of gradient in infinite-dimensional spaces should be sufficiently general to fit with various classes of nonsmooth functionals. For easy of exposition, in this chapter we decided to focus our attention on the case of *geodesically convex* (or, more generally, λ -convex) functionals (we refer to [9] for more general results). Geodesics in $\mathcal{P}_2(\mathbb{R}^d)$ play a crucial role and their characterization is briefly discussed in Section 2.3. Section 3 is thus devoted to the analysis of convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$ and to some particularly important examples, discovered by McCann [66].

Having at our disposal a nice Hilbertian structure at the level of each tangent space and a significant notion of convexity, it is natural to develop a *subdifferential theory* modeled on the well-known linear one. We deal with this program in Section 4: first of all we define the (Fréchet) subdifferential $\partial\phi(\mu)$ of ϕ at a measure μ . Even if it is a multivalued map, it is possible to perform a natural *minimal selection* $\partial\phi^\circ(\mu)$ among its values, which enjoys nice features and always belongs to the tangent space $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$. Sections 4.2–4.4 present the basic calculus properties of the subdifferential: they precisely reproduce the analogous ones of the linear framework and justify the interest for this notion. Section 4.5 contains the main characterizations of the subdifferential of the most relevant functionals (internal, potential and interaction energies, and the negative squared Wasserstein distance).

Combining all these notions, we end up with the rigorous definition of the gradient flow of a functional ϕ in Section 5: it always has the structure of the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \quad \text{in } \mathcal{D}'(\mathbb{R}^d \times (0, T)), \quad (0.13)$$

which defines the *velocity* of μ_t , coupled with the *nonlinear condition*

$$\mathbf{v}_t = -\partial^\circ \phi(\mu_t) \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, T), \quad (0.14)$$

linking $\mathbf{v}_t$ to μ_t through the functional ϕ . When ϕ has the structure of (0.6) and $\mu_t = \rho_t \gamma$, (0.14) is equivalent (in a suitable weak sense) to

$$\mathbf{v}_t = -\nabla F'(\rho_t). \quad (0.15)$$

The remaining part of the section is devoted to study the *main properties of the gradient flows*, obtained independently from the existence issue, i.e. directly from the definition. We conclude the section providing an answer to the second question we raised before, i.e., the *construction of the gradient flow by means of the variational approximation scheme*.

Even in this case (λ -geodesic) convexity plays a crucial role and we are able to obtain the same well-known results of the theory in flat linear spaces. Here we only mention the generation of a contracting and regularizing semigroup satisfying, when $\lambda > 0$, nice asymptotic convergence estimates. In comparison with other papers ([29,76], for the porous medium equation on Riemannian manifolds), where similar goals are pursued, our approach is totally independent of the specific form of the functional ϕ and of the PDE that it induces: it is ultimately based on the one hand on monotonicity inequalities (ensured by the λ -convexity of ϕ), and on the other hand on (0.12), whose validity is a purely geometrical fact. Furthermore, as shown in [9], it extends also to the case when $\mathbb{R}^d$ is replaced by a separable Hilbert space and/or singular (e.g., concentrated) measures are allowed.

The last section illustrates our main examples and applications. A particular emphasis is devoted to the linear Fokker–Planck equation (0.4) associated to a convex potential V with arbitrary growth at infinity: as showed by Otto, it is the gradient flow in $\mathcal{P}_2(\mathbb{R}^d)$ of the relative entropy functional

$$\phi(\mu) := \int_{\mathbb{R}^d} \rho(x) \log \rho(x) d\gamma(x), \quad \rho = \frac{d\mu}{d\gamma}, \quad (0.16)$$

with respect to the invariant measure $\gamma := e^{-V} \mathcal{L}^d$. In this case the Wasserstein approach provides a linear semigroup in the space of measures (a Dirac mass concentrated in a point where the potential is finite is always allowed as an initial datum), which easily gives nice representation formulae for the solution. The restriction of the semigroup on absolutely continuous measures w.r.t. γ coincides with the Markov semigroup generated by the natural Dirichlet form associated to γ .

Applications to the case of nonlinear diffusion equations and to more complicated differential–integral equations are also considered.

Notation

$B_r(x)$	open ball of radius r centered at x in a metric space
$\mathcal{B}(X)$	Borel sets in a separable metric space X
$C_b^0(X)$	space of continuous and bounded real functions defined on X
$C_c^\infty(\mathbb{R}^d)$	space of smooth real functions with compact support in $\mathbb{R}^d$
$\mathcal{P}(X)$	probability measures in a separable metric space X
$\mathcal{P}_2(X)$	probability measures with finite quadratic moment, see (1.3)
$\mathcal{L}^d$	the Lebesgue measure in $\mathbb{R}^d$

$\mathcal{P}_2^a(\mathbb{R}^d)$	measures in $\mathcal{P}_2(\mathbb{R}^d)$ absolutely continuous w.r.t. $\mathcal{L}^d$
$L^p(\mu; \mathbb{R}^d)$	L^p space of μ -measurable $\mathbb{R}^d$ -valued maps
$\text{supp } \mu$	support of μ , see (1.1)
$\mathbf{r}_\# \mu$	push-forward of μ through $\mathbf{r}$, see (1.4)
π^i	projection operators on a product space $\mathbf{X}$, see (1.8)
$\Gamma(\mu^1, \mu^2)$	2-plans with given marginals μ^1, μ^2
$\Gamma_o(\mu^1, \mu^2)$	optimal 2-plans with given marginals μ^1, μ^2
$W_2(\mu, \nu)$	Wasserstein distance between μ and ν , see (2.6)
$\mathbf{i}$	identity map
$\mathbf{t}_\mu^\nu$	optimal transport map between μ and ν given by Theorem 2.3
$\text{Tan}_{\mu_i} \mathcal{P}_2(\mathbb{R}^d)$	tangent bundle to $\mathcal{P}_2(\mathbb{R}^d)$, see (2.42)
$\mu_t^{1 \rightarrow 2}$	geodesic curve connecting μ^1 to μ^2 , see (3.1)
$ u' (t)$	metric derivative of $u : (a, b) \rightarrow E$, see (2.2)
$AC^p((a, b); E)$	absolutely continuous $u : (a, b) \rightarrow E$ with $ u' \in L^p(a, b)$, see (2.3)
$D(\phi)$	proper domain of a functional ϕ , see (4.1)
$\text{Lip}(\phi, A)$	Lipschitz constant of the function ϕ in the set A
$\partial \phi(v)$	Fréchet subdifferential of ϕ in Hilbert (4.2) or Wasserstein spaces, see Definition 4.1 and (4.20)
$ \partial \phi (v)$	metric slope of ϕ , see (4.4) and (4.29)
$\partial^\circ \phi(\mu)$	minimal selection in the subdifferential, see Lemma 4.10
$\bar{M}_\tau(t)$	piecewise constant interpolation of M_τ^n , see (5.54)
$MM(\Phi; u_0)$	minimizing movement of ϕ , see the definition before (5.55)

1. Notation and measure-theoretic results

In this section we recall the main notation used in this chapter and some basic measure-theoretic terminology and results. Given a separable metric space (X, d) , we denote by $\mathcal{P}(X)$ the set of probability measures $\mu : \mathcal{B}(X) \rightarrow [0, 1]$, where $\mathcal{B}(X)$ is the Borel σ -algebra. The support of $\mu \in \mathcal{P}(X)$ is the closed set

$$\text{supp}(\mu) := \{x \in X : \mu(B_r(x)) > 0 \ \forall r > 0\}. \quad (1.1)$$

When X is a Borel subset of an euclidean space $\mathbb{R}^d$, we set

$$m_2(\mu) := \int_X |x|^2 d\mu,$$

we often make the identification

$$\mathcal{P}(X) = \{\mu \in \mathcal{P}(\mathbb{R}^d) : \mu(\mathbb{R}^d \setminus X) = 0\}, \quad (1.2)$$

and we denote by $\mathcal{P}_2(X)$ the subspace of $\mathcal{P}(X)$ made by measures with finite quadratic moment:

$$\mathcal{P}_2(X) := \{\mu \in \mathcal{P}(X) : m_2(\mu) < \infty\}. \quad (1.3)$$

We denote by $\mathcal{L}^d$ the Lebesgue measure in $\mathbb{R}^d$ and set

$$\mathcal{P}_2^a(X) := \{\mu \in \mathcal{P}_2(X) : \mu \ll \mathcal{L}^d\},$$

whenever $X \in \mathcal{B}(\mathbb{R}^d)$.

1.1. Transport maps and transport plans

If $\mu \in \mathcal{P}(X_1)$, and $\mathbf{r} : X_1 \rightarrow X_2$ is a Borel (or, more generally, μ -measurable) map, we denote by $\mathbf{r}_\# \mu \in \mathcal{P}(X_2)$ the *push-forward of μ through $\mathbf{r}$* , defined by

$$\mathbf{r}_\# \mu(B) := \mu(\mathbf{r}^{-1}(B)) \quad \forall B \in \mathcal{B}(X_2). \quad (1.4)$$

More generally, we have

$$\int_{X_1} f(\mathbf{r}(x)) d\mu(x) = \int_{X_2} f(y) d\mathbf{r}_\# \mu(y) \quad (1.5)$$

for every bounded (or $\mathbf{r}_\# \mu$ -integrable) Borel function $f : X_2 \rightarrow \mathbb{R}$. It is easy to check that

$$v \ll \mu \implies \mathbf{r}_\# v \ll \mathbf{r}_\# \mu \quad \forall \mu, v \in \mathcal{P}(X_1). \quad (1.6)$$

Notice also the natural composition rule

$$(\mathbf{r} \circ \mathbf{s})_\# \mu = \mathbf{r}_\# (\mathbf{s}_\# \mu) \quad \text{where } \mathbf{s} : X_1 \rightarrow X_2, \mathbf{r} : X_2 \rightarrow X_3, \mu \in \mathcal{P}(X_1). \quad (1.7)$$

We denote by π^i , $i = 1, 2$, the projection operators defined on a product space $\mathbf{X} := X_1 \times X_2$, defined by

$$\pi^1 : (x_1, x_2) \mapsto x_1 \in X_1, \quad \pi^2 : (x_1, x_2) \mapsto x_2 \in X_2. \quad (1.8)$$

If $\mathbf{X}$ is endowed with the canonical product metric and the Borel σ -algebra and $\mu \in \mathcal{P}(\mathbf{X})$, the *marginals* of μ are the probability measures

$$\mu^i := \pi_{\#}^i \mu \in \mathcal{P}(X_i), \quad i = 1, 2. \quad (1.9)$$

Given $\mu^1 \in \mathcal{P}(X_1)$ and $\mu^2 \in \mathcal{P}(X_2)$ the class $\Gamma(\mu^1, \mu^2)$ of transport plans between μ^1 and μ^2 is defined by

$$\Gamma(\mu^1, \mu^2) := \{\mu \in \mathcal{P}(X_1 \times X_2) : \pi_{\#}^i \mu = \mu^i, i = 1, 2\}. \quad (1.10)$$

Notice also that

$$\Gamma(\mu^1, \mu^2) = \{\mu^1 \times \mu^2\} \quad \text{if either } \mu^1 \text{ or } \mu^2 \text{ is a Dirac mass.} \quad (1.11)$$

To each couple of measures $\mu^1 \in \mathcal{P}(X_1)$, $\mu^2 = \mathbf{r}_\# \mu^1 \in \mathcal{P}(X_2)$ linked by a Borel transport map $\mathbf{r}: X_1 \rightarrow X_2$ we can associate the transport plan

$$\mu := (\mathbf{i} \times \mathbf{r})_\# \mu^1 \in \Gamma(\mu^1, \mu^2), \quad \mathbf{i} \text{ being the identity map on } X_1. \quad (1.12)$$

If μ is representable as in (1.12) then we say that μ is *induced* by $\mathbf{r}$. Each transport plan μ concentrated on a μ -measurable graph in $X_1 \times X_2$ admits the representation (1.12) for some μ^1 -measurable map $\mathbf{r}$, which therefore transports μ^1 to μ^2 (see, e.g., [7]).

1.2. Narrow convergence

Conformally to the probabilistic terminology, we say that a sequence $(\mu_n) \subset \mathcal{P}(X)$ is *narrowly convergent* to $\mu \in \mathcal{P}(X)$ as $n \rightarrow \infty$ if

$$\lim_{n \rightarrow \infty} \int_X f(x) d\mu_n(x) = \int_X f(x) d\mu(x) \quad (1.13)$$

for every function $f \in C_b^0(X)$, the space of continuous and bounded real functions defined on X .

THEOREM 1.1 ([39], III-59). *If a set $\mathcal{K} \subset \mathcal{P}(X)$ is tight, i.e.,*

$$\forall \varepsilon > 0 \exists K_\varepsilon \text{ compact in } X \text{ such that } \mu(X \setminus K_\varepsilon) \leq \varepsilon \quad \forall \mu \in \mathcal{K}, \quad (1.14)$$

then $\mathcal{K}$ is relatively compact in $\mathcal{P}(X)$.

When one needs to pass to the limit in expressions like (1.13) w.r.t. *unbounded or lower semicontinuous* functions f , the following two properties are quite useful. The first one is a lower semicontinuity property,

$$\liminf_{n \rightarrow \infty} \int_X g(x) d\mu_n(x) \geq \int_X g(x) d\mu(x) \quad (1.15)$$

for every sequence $(\mu_n) \subset \mathcal{P}(X)$ narrowly convergent to μ and any l.s.c. function $g: X \rightarrow (-\infty, +\infty]$ bounded from below: it follows easily by a monotone approximation argument of g by continuous and bounded functions. Changing g in $-g$ one gets the corresponding “lim sup” inequality for upper semicontinuous functions bounded from above. In particular, choosing as g the characteristic functions of open and closed subset of X , we obtain

$$\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G) \quad \forall G \text{ open in } X, \quad (1.16)$$

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F) \quad \forall F \text{ closed in } X. \quad (1.17)$$

The statement of the second property requires the following definitions: we say that a Borel function $g : X \rightarrow [0, +\infty]$ is *uniformly integrable* w.r.t. a given set $\mathcal{K} \subset \mathcal{P}(X)$ if

$$\lim_{k \rightarrow \infty} \int_{\{x: g(x) \geq k\}} g(x) d\mu(x) = 0 \quad \text{uniformly w.r.t. } \mu \in \mathcal{K}. \quad (1.18)$$

In the particular case of $g(x) := d(x, \bar{x})^p$, for some (and thus any) $\bar{x} \in X$ and a given $p > 0$, i.e., if

$$\lim_{k \rightarrow \infty} \int_{X \setminus B_k(\bar{x})} d^p(\bar{x}, x) d\mu(x) = 0 \quad \text{uniformly w.r.t. } \mu \in \mathcal{K}, \quad (1.19)$$

we say that the set $\mathcal{K} \subset \mathcal{P}(X)$ has *uniformly integrable p -moments*. The following lemma (see, for instance, Lemma 5.1.7 of [9] for its proof) provides a characterization of p -uniformly integrable families, extending the validity of (1.13) to unbounded but with p -growth functions, i.e., functions $f : X \rightarrow \mathbb{R}$ such that

$$|f(x)| \leq A + B d^p(\bar{x}, x) \quad \forall x \in X, \quad (1.20)$$

for some $A, B \geq 0$ and $\bar{x} \in X$.

LEMMA 1.2. *Let $(\mu_n) \subset \mathcal{P}(X)$ be narrowly convergent to $\mu \in \mathcal{P}(X)$. If $f : X \rightarrow \mathbb{R}$ is continuous, $g : X \rightarrow (-\infty, +\infty]$ is lower semicontinuous, and $|f|$ and g^- are uniformly integrable w.r.t. the set $\{\mu_n\}_{n \in \mathbb{N}}$, then*

$$\liminf_{n \rightarrow \infty} \int_X g(x) d\mu_n(x) \geq \int_X g(x) d\mu(x) > -\infty, \quad (1.21a)$$

$$\lim_{n \rightarrow \infty} \int_X f(x) d\mu_n(x) = \int_X f(x) d\mu(x). \quad (1.21b)$$

Conversely, if $f : X \rightarrow [0, \infty)$ is continuous, μ_n -integrable, and

$$\limsup_{n \rightarrow \infty} \int_X f(x) d\mu_n(x) \leq \int_X f(x) d\mu(x) < +\infty, \quad (1.22)$$

then f is uniformly integrable w.r.t. $\{\mu_n\}_{n \in \mathbb{N}}$.

In particular, a family $\{\mu_n\}_{n \in \mathbb{N}} \subset \mathcal{P}(X)$ has uniformly integrable p -moments iff (1.21b) holds for every continuous function $f : X \rightarrow \mathbb{R}$ with p -growth.

1.3. The change of variables formula

Let $\mathbf{r} : A \subset \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a Borel function, with A open. Then, denoting by $\Sigma_{\mathbf{r}} = D(\nabla \mathbf{r})$ the Borel set where $\mathbf{r}$ is differentiable, there is a sequence of sets $\Sigma_n \uparrow \Sigma_{\mathbf{r}}$ such that $\mathbf{r}|_{\Sigma_n}$ is a Lipschitz function for any n (see [45], Section 3.1.8). Therefore the well-known area

formula for Lipschitz maps (see, for instance, [44,45]) extends to this general class of maps and reads as follows:

$$\int_{\Sigma_{\mathbf{r}}} h(x) |\det \nabla \mathbf{r}|(x) dx = \int_{\mathbb{R}^d} \sum_{x \in \Sigma_{\mathbf{r}} \cap \mathbf{r}^{-1}(y)} h(x) dy \quad (1.23)$$

for any Borel function $h : \mathbb{R}^d \rightarrow [0, +\infty]$. This formula leads to a simple rule for computing the density of the push-forward of measures absolutely continuous w.r.t. $\mathcal{L}^d$.

LEMMA 1.3 (Density of the push-forward). *Let $\rho \in L^1(\mathbb{R}^d)$ be a nonnegative function and assume that there exists a Borel set $\Sigma \subset \Sigma_{\mathbf{r}}$ such that $\mathbf{r}|_{\Sigma}$ is injective and the difference $\{\rho > 0\} \setminus \Sigma$ is $\mathcal{L}^d$ -negligible. Then $\mathbf{r}_{\#}(\rho \mathcal{L}^d) \ll \mathcal{L}^d$ if and only if $|\det \nabla \mathbf{r}| > 0$ $\mathcal{L}^d$ -a.e. on Σ and in this case*

$$\mathbf{r}_{\#}(\rho \mathcal{L}^d) = \frac{\rho}{|\det \nabla \mathbf{r}|} \circ \mathbf{r}^{-1} \Big|_{\mathbf{r}(\Sigma)} \mathcal{L}^d.$$

PROOF. If $|\det \nabla \mathbf{r}| > 0$ $\mathcal{L}^d$ -a.e. on Σ we can put $h = \rho \chi_{\mathbf{r}^{-1}(B) \cap \Sigma} / |\det \nabla \mathbf{r}|$ in (1.23), with $B \in \mathcal{B}(\mathbb{R}^d)$, to obtain

$$\int_{\mathbf{r}^{-1}(B)} \rho dx = \int_{\mathbf{r}^{-1}(B) \cap \Sigma} \rho dx = \int_{B \cap \mathbf{r}(\Sigma)} \frac{\rho(\mathbf{r}^{-1}(y))}{|\det \nabla \mathbf{r}(\mathbf{r}^{-1}(y))|} dy.$$

Conversely, if there is a Borel set $B \subset \Sigma$ with $\mathcal{L}^d(B) > 0$ and $|\det \nabla \mathbf{r}| = 0$ on B , the area formula gives $\mathcal{L}^d(\mathbf{r}(B)) = 0$. On the other hand,

$$\mathbf{r}_{\#}(\rho \mathcal{L}^d)(\mathbf{r}(B)) = \int_{\mathbf{r}^{-1}(\mathbf{r}(B))} \rho dx > 0$$

because at $\mathcal{L}^d$ -a.e. $x \in B$ we have $\rho(x) > 0$. Hence, $\mathbf{r}_{\#}(\rho \mathcal{L}^d)$ is not absolutely continuous with respect to $\mathcal{L}^d$. $\square$

By applying the area formula again, we obtain the rule for computing integrals of the densities

$$\int_{\mathbb{R}^d} F\left(\frac{\mathbf{r}_{\#}(\rho \mathcal{L}^d)}{\mathcal{L}^d}\right) dx = \int_{\mathbb{R}^d} F\left(\frac{\rho}{|\det \nabla \mathbf{r}|}\right) |\det \nabla \mathbf{r}| dx \quad (1.24)$$

for any Borel function $F : [0, +\infty) \rightarrow [0, +\infty]$ with $F(0) = 0$. Notice that in this formula the set $\Sigma_{\mathbf{r}}$ does not appear anymore (due to the fact that $F(0) = 0$ and $\rho = 0$ out of Σ), so it holds provided $\mathbf{r}$ is differentiable $\rho \mathcal{L}^d$ -a.e., it is $\rho \mathcal{L}^d$ -essentially injective (i.e., there exists a Borel set Σ such that $\mathbf{r}|_{\Sigma}$ is injective and $\rho = 0$ $\mathcal{L}^d$ -a.e. out of Σ) and $|\det \nabla \mathbf{r}| > 0$ $\rho \mathcal{L}^d$ -a.e. in $\mathbb{R}^d$.

We will apply mostly these formulas when $\mathbf{r}$ is the gradient of a convex function $g : \Omega \rightarrow \mathbb{R}$, Ω being an open subset of $\mathbb{R}^d$. In this specific case it is well known that

the (multivalued) subdifferential $\partial g(x)$ of g (we will recall its definition at the beginning of Section 4) is nonempty for every $x \in \Omega$ and it is reduced to a single point $\nabla g(x)$ when g is differentiable at x : this happens for $\mathcal{L}^d$ -a.e. $x \in \Omega$.

In the following result (see, for instance, [4,44]) we are considering an arbitrary Borel selection $\mathbf{r} : \Omega \rightarrow \mathbb{R}^d$ such that

$$\mathbf{r}(x) \in \partial g(x) \quad \text{for every } x \in \Omega. \quad (1.25)$$

THEOREM 1.4 (Aleksandrov). *Let $\Omega \subset \mathbb{R}^d$ be a convex open set and let $g : \Omega \rightarrow \mathbb{R}$ be a convex function. Then g is a locally Lipschitz function, (every extension $\mathbf{r}$ satisfying (1.25) of) ∇g is differentiable at $\mathcal{L}^d$ -a.e. point of Ω , its gradient $\nabla^2 g(x)$ is a symmetric matrix, and g has the second-order Taylor expansion*

$$g(y) = g(x) + \langle \nabla g(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 g(x), y - x \rangle + o(|y - x|^2) \quad \text{as } y \rightarrow x \quad (1.26)$$

for $\mathcal{L}^d$ -a.e. $x \in \Omega$.

Notice that ∇g is also monotone

$$\langle \nabla g(x_1) - \nabla g(x_2), x_1 - x_2 \rangle \geq 0, \quad x_1, x_2 \in D(\nabla g),$$

and that the above inequality is strict if g is *strictly* convex: in this case, it is immediate to check that ∇g is injective on $D(\nabla g)$, and that $|\det \nabla^2 g| > 0$ on the differentiability set of ∇g if g is *uniformly* convex.

2. Metric and differentiable structure of the Wasserstein space

In this section we look at $\mathcal{P}_2(\mathbb{R}^d)$ first from the metric and then from the differentiable viewpoints.

2.1. Absolutely continuous maps and metric derivative

Let (E, d) be a metric space.

DEFINITION 2.1 (Absolutely continuous curves). Let $I \subset \mathbb{R}$ be an interval and let $u : I \rightarrow E$. We say that u is absolutely continuous if there exists $m \in L^1(I)$ such that

$$d(u(s), u(t)) \leq \int_s^t m(\tau) d\tau \quad \forall s, t \in I, s \leq t. \quad (2.1)$$

Any absolutely continuous curve is obviously uniformly continuous, and therefore it can be uniquely extended to the closure of I . It is not difficult to show (see, for instance, Theorem 1.1.2 in [9] or [11]) that the *metric derivative*

$$|u'| (t) := \lim_{h \rightarrow 0} \frac{d(u(t+h), u(t))}{|h|} \quad (2.2)$$

exists at $\mathcal{L}^1$ -a.e. $t \in I$ for any absolutely continuous curve $u(t)$. Furthermore, $|u'| \in L^1(I)$ and is the minimal m fulfilling (2.1) (i.e., $|u'|$ fulfills (2.1) and $m \geq |u'|$ $\mathcal{L}^1$ -a.e. in I for any m with this property). For $p \in [1, +\infty]$ we also set

$$\begin{aligned} AC^p(I; E) \\ := \{u : I \rightarrow E : u \text{ is absolutely continuous and } |u'| \in L^p(I)\}. \end{aligned} \quad (2.3)$$

2.2. The quadratic optimal transport problem

Let X, Y be complete and separable metric spaces and let $c : X \times Y \rightarrow [0, +\infty]$ be a Borel cost function. Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ the optimal transport problem, in Monge's formulation, is given by

$$\inf \left\{ \int_X c(x, \mathbf{t}(x)) d\mu(x) : \mathbf{t}_\# \mu = \nu \right\}. \quad (2.4)$$

This problem can be ill posed because sometimes there is no transport map $\mathbf{t}$ such that $\mathbf{t}_\# \mu = \nu$ (this happens for instance when μ is a Dirac mass and ν is not a Dirac mass). Kantorovich's formulation

$$\min \left\{ \int_{X \times Y} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu) \right\} \quad (2.5)$$

circumvents this problem (as $\mu \times \nu \in \Gamma(\mu, \nu)$). The existence of an optimal transport plan, when c is l.s.c., is provided by (1.15) and by Theorem 1.1, taking into account that $\Gamma(\mu, \nu)$ is tight (this follows easily by the fact that the marginals of the measures in $\Gamma(\mu, \nu)$ are fixed, and by the fact that according to Ulam's theorem any finite measure in a complete and separable metric space is tight, see also Chapter 6 in [9] for more general formulations).

The problem (2.5) is truly a weak formulation of (2.4) in the following sense: if c is bounded and continuous, and if μ has no atom, then the “min” in (2.5) is equal to the “inf” in (2.4), see [7, 47]. This result can also be extended to classes of unbounded cost functions, see [79].

In the sequel we consider the case when $X = Y$ and $c(x, y) = d^2(x, y)$, where d is the distance in X , and denote by $\Gamma_o(\mu, \nu)$ the *optimal plans* in (2.5) corresponding to this choice of the cost function. In this case we use the minimum value to define the Kantorovich–Rubinstein–Wasserstein distance

$$W_2(\mu, \nu) := \left(\int_{X \times X} d^2(x, y) d\gamma \right)^{1/2}, \quad \gamma \in \Gamma_o(\mu, \nu). \quad (2.6)$$

THEOREM 2.2. *Let X be a complete and separable metric space. Then W_2 defines a distance in $\mathcal{P}_2(X)$ and $\mathcal{P}_2(X)$, endowed with this distance, is a complete and separable metric space. Furthermore, for a given sequence $(\mu_n) \subset \mathcal{P}_2(X)$ we have*

$$\begin{aligned} \lim_{n \rightarrow \infty} W_2(\mu_n, \mu) = 0 \\ \iff \begin{cases} \mu_n \text{ narrowly converge to } \mu, \\ (\mu_n) \text{ has uniformly integrable 2-moments.} \end{cases} \end{aligned} \quad (2.7)$$

PROOF. We just prove that W_2 is a distance. The complete statement is proved for instance in Proposition 7.1.5 of [9] or, in the locally compact case, in [86].

Let $\mu, \nu, \sigma \in \mathcal{P}_2(X)$ and let $\gamma \in \Gamma_0(\mu, \nu)$ and $\eta \in \Gamma_0(\nu, \sigma)$. General results of probability theory (see the above mentioned references) ensure the existence of $\lambda \in \mathcal{P}(X \times X \times X)$ such that

$$(\pi^1, \pi^2)_\# \lambda = \gamma, \quad (\pi^2, \pi^3)_\# \lambda = \eta.$$

Then, as

$$\pi_\#^1(\pi^1, \pi^3)_\# \lambda = \pi_\#^1 \lambda = \pi_\#^1 \gamma = \mu, \quad \pi_\#^2(\pi^1, \pi^3)_\# \lambda = \pi_\#^2 \lambda = \pi_\#^2 \eta = \sigma,$$

we obtain that $(\pi^1, \pi^3)_\# \lambda \in \Gamma(\mu, \sigma)$, hence

$$W_2(\mu, \sigma) \leq \left(\int_{X \times X} d^2(x_1, x_3) d(\pi^1, \pi^3)_\# \lambda \right)^{1/2} = \|d(x_1, x_3)\|_{L^2(\lambda)}.$$

As $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$ and

$$\begin{aligned} \|d(x_1, x_2)\|_{L^2(\lambda)} &= \|d(x_1, x_2)\|_{L^2(\gamma)} = W_2(\mu, \nu), \\ \|d(x_2, x_3)\|_{L^2(\lambda)} &= \|d(x_2, x_3)\|_{L^2(\eta)} = W_2(\nu, \sigma), \end{aligned}$$

the triangle inequality $W_2(\mu, \sigma) \leq W_2(\mu, \nu) + W_2(\nu, \sigma)$ follows by the standard triangle inequality in $L^2(\lambda)$. $\square$

In the Euclidean case $X = \mathbb{R}^d$, notice that, thanks to Lemma 1.2, the uniform integrability of $|x|^2$ with respect to $\{\mu_n\}_{n \in \mathbb{N}}$ is equivalent, assuming the narrow convergence of μ_n to μ , to the convergence of $m_2(\mu_n)$ to $m_2(\mu)$. Both conditions in the right-hand side of (2.7) can be summarized, still thanks to the same lemma, by saying that (1.21b) holds for any continuous function f with at most quadratic growth.

Working with Monge's formulation the proof above is technically easier, as an admissible transport map between μ and σ can be obtained just composing transport maps between μ and ν with transport maps between ν and σ . However, in order to give a complete proof one needs to know either that optimal plans are induced by maps, or that the infimum in Monge's formulation coincides with the minimum in Kantorovich's one, and none of these results is trivial, even in Euclidean spaces.

Although in many situations that we consider in this chapter the optimal plans are induced by maps, still the Kantorovich formulation of the optimal transport problem is quite useful to provide estimates *from above* on W_2 . For instance,

$$W_2^2(\mu, \nu) \leq \int_X d^2(\mathbf{t}(x), \mathbf{s}(x)) d\sigma(x) \quad \text{whenever } \mathbf{t}_\# \sigma = \mu, \mathbf{s}_\# \sigma = \nu. \quad (2.8)$$

This follows by the fact that $(\mathbf{t}, \mathbf{s})_\# \sigma \in \Gamma(\mu, \nu)$ and by the identity

$$\int_X d^2(\mathbf{t}(x), \mathbf{s}(x)) d\sigma(x) = \int_{X \times X} d^2(x, y) d(\mathbf{t}, \mathbf{s})_\# \sigma.$$

2.3. Geodesics in $\mathcal{P}_2(\mathbb{R}^d)$

Let (E, d) be a metric space. Recall that a constant speed geodesic $\gamma : [0, T] \rightarrow E$ is a map satisfying

$$d(\gamma(s), \gamma(t)) = \frac{(t-s)}{T} d(\gamma(0), \gamma(T)) \quad \text{whenever } 0 \leq s \leq t \leq T.$$

Actually only the inequality $d(\gamma(s), \gamma(t)) \leq T^{-1}(t-s)d(\gamma(0), \gamma(T))$ needs to be checked for all $0 \leq s \leq t \leq T$. Indeed, if the strict inequality occurs for some $s < t$, then the triangle inequality provides

$$\begin{aligned} d(\gamma(0), \gamma(T)) &\leq d(\gamma(0), \gamma(s)) + d(\gamma(s), \gamma(t)) + d(\gamma(t), \gamma(T)) \\ &< \frac{1}{T}(s + (t-s) + (T-t))d(\gamma(0), \gamma(T)) \\ &= d(\gamma(0), \gamma(T)), \end{aligned}$$

a contradiction.

Using this elementary fact one can show that, for any choice of $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, and $\gamma \in \Gamma_0(\mu, \nu)$, the map

$$\mu_t := ((1-t)\pi^1 + t\pi^2)_\# \gamma, \quad t \in [0, 1], \quad (2.9)$$

is a constant speed geodesic. Indeed,

$$\gamma_{st} := (((1-s)\pi^1 + s\pi^2), ((1-t)\pi^1 + t\pi^2))_\# \gamma \in \Gamma(\mu_s, \mu_t)$$

and this plan provides the estimate

$$W_2(\mu_s, \mu_t) \leq (t-s)W_2(\mu, \nu), \quad (2.10)$$

as

$$\begin{aligned} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^2 d\gamma_{st} &= \int_{\mathbb{R}^d \times \mathbb{R}^d} |(1-s)x_1 + sx_2 - (1-t)x_1 - tx_2|^2 d\gamma \\ &= (s-t)^2 \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^2 d\gamma. \end{aligned}$$

It has been proved in Theorem 7.2.2 of [9] that *any* constant speed geodesic joining μ to ν can be built in this way. We discuss additional regularity properties of the geodesics in the next section. Here we just mention that, in the case when γ is induced by a transport map $\mathbf{t}$ (i.e., $\gamma = (\mathbf{i}, \mathbf{t})_{\#}\mu$), then (2.9) reduces to

$$\mu_t = ((1-t)\mathbf{i} + t\mathbf{t})_{\#}\mu, \quad t \in [0, 1]. \quad (2.11)$$

2.4. Existence of optimal transport maps

The following basic result of [20,48,60] provides existence and uniqueness of the optimal transport map in the case when the initial measure μ belongs to $\mathcal{P}_2^a(\mathbb{R}^d)$.

THEOREM 2.3 (Existence and uniqueness of optimal transport maps). *For any $\mu \in \mathcal{P}_2^a(\mathbb{R}^d)$, $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ Kantorovich's optimal transport problem (2.5) with $c(x, y) = |x - y|^2$ has a unique solution γ . Moreover:*

(i) γ is induced by a transport map $\mathbf{t}$, i.e., $\gamma = (\mathbf{i}, \mathbf{t})_{\#}\mu$. In particular $\mathbf{t}$ is the unique solution of Monge's optimal transport problem (2.4).

(ii) The map $\mathbf{t}$ coincides μ -a.e. with the gradient of a convex function $\varphi: \mathbb{R}^d \rightarrow (-\infty, +\infty]$, whose finiteness domain $D(\varphi)$ has nonempty interior and satisfies

$$\mu(\mathbb{R}^d \setminus D(\varphi)) = \mu(\mathbb{R}^d \setminus D(\nabla\varphi)) = 0. \quad (2.12)$$

(iii) If $\nu = \rho' \mathcal{L}^d \in \mathcal{P}_2^a(\mathbb{R}^d)$ as well, and $\mathbf{s}$ is the optimal transport map between ν and μ , then

$$\mathbf{s} \circ \mathbf{t} = \mathbf{i} \quad \mu\text{-a.e. in } \mathbb{R}^d \quad \text{and} \quad \mathbf{t} \circ \mathbf{s} = \mathbf{i} \quad \nu\text{-a.e. in } \mathbb{R}^d.$$

In particular, $\mathbf{t}$ is μ -essentially injective, i.e., there exists a μ -negligible set $\mathcal{N} \subset \mathbb{R}^d$ such that, setting $\Omega = \mathbb{R}^d \setminus \mathcal{N}$, $\mathbf{t}|_{\Omega}$ is injective. Finally,

$$\rho' := \frac{\rho}{\det \nabla^2 \varphi} \circ (\mathbf{t}|_{\Omega})^{-1} \quad \nu\text{-a.e. in } \mathbb{R}^d.$$

PROOF. We are presenting here the proof of the last statement (iii). Since $(\mathbf{i}, \mathbf{t})_{\#}\mu$ and $(\mathbf{s}, \mathbf{i})_{\#}\nu$ are both optimal plans between μ and ν , they coincide. Testing this identity

between plans on $|\mathbf{s}(\mathbf{t}(x)) - x|$ (resp. $|\mathbf{t}(\mathbf{s}(y)) - y|$) we obtain that $\mathbf{s} \circ \mathbf{t} = \mathbf{i}$ μ -a.e. in $\mathbb{R}^d$ (resp. $\mathbf{t} \circ \mathbf{s} = \mathbf{i}$ ν -a.e. in $\mathbb{R}^d$),

$$\begin{aligned} \int_{\mathbb{R}^d} |x - \mathbf{s}(\mathbf{t}(x))| d\mu(x) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - \mathbf{s}(y)| d(\mathbf{i}, \mathbf{t})_{\#} \mu \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - \mathbf{s}(y)| d(\mathbf{s}, \mathbf{i})_{\#} \nu \\ &= \int_{\mathbb{R}^d} |\mathbf{s}(y) - \mathbf{s}(y)| d\nu(y) = 0. \end{aligned}$$

The formula for the density of ν with respect to $\mathcal{L}^d$ follows by Lemma 1.3, taking into account the μ -essential injectivity of $\mathbf{t}$. $\square$

In the following we shall denote by $\mathbf{t}_{\mu}^{\nu}$ the unique optimal map given by Theorem 2.3. Notice that $\mathbf{t} = \nabla\varphi$ is uniquely determined only μ -a.e., hence φ is not uniquely determined, not even up to additive constants, unless $\mu = \rho\mathcal{L}^d$ with $\rho > 0$ $\mathcal{L}^d$ -a.e. in $\mathbb{R}^d$. However, the existence proof (at least the one achieved through a duality argument), yields some “canonical” φ , given by the duality formula

$$\varphi(x) = \sup_{y \in \text{supp } \nu} \langle x, y \rangle - \psi(y), \quad x \in \mathbb{R}^d, \quad (2.13)$$

for a suitable function $\psi : \text{supp } \nu \rightarrow (-\infty, +\infty]$. This explicit expression is sometimes technically useful: for instance, it shows that when $\text{supp } \nu$ is bounded we can always find a globally convex and Lipschitz map φ whose gradient is the optimal transport map.

The following result shows that optimal maps along geodesics enjoy nicer properties (see also [17]).

THEOREM 2.4 (Regularity in the interior of geodesics). *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and let*

$$\mu_t := ((1-t)\pi^1 + t\pi^2)_{\#} \gamma$$

be a constant speed geodesic induced by $\gamma \in \Gamma_0(\mu, \nu)$. Then the following properties hold:

- (i) *For any $t \in [0, 1)$ there exists a unique optimal plan between μ_t and μ , and this plan is induced by a map $\mathbf{s}_t$ with Lipschitz constant less than $1/(1-t)$.*
- (ii) *If $\mu = \rho\mathcal{L}^d \in \mathcal{P}_2^a(\mathbb{R}^d)$ then $\mu_t \in \mathcal{P}_2^a(\mathbb{R}^d)$ for all $t \in [0, 1)$.*

PROOF. (i) The necessary optimality conditions at the level of plans (see, for instance, Section 6.2.3 of [9], or [86]) imply that the support of γ is contained in the graph

$$\{(x, y) : y \in \Gamma(x)\}$$

of a monotone operator $\Gamma(x)$. On the other hand, the same argument used in the proof of (2.10) shows that the plan $\gamma_t := (\pi^1, (1-t)\pi^1 + t\pi^2)_{\#} \gamma$ is optimal between μ and μ_t .

The support of γ_t is contained in the graph of the monotone operator $(1-t)I + t\Gamma$, whose inverse

$$\Gamma^{-1}(y) := \{x \in \mathbb{R}^d : y \in \Gamma(x)\}$$

is single-valued and $1/(1-t)$ -Lipschitz continuous. Therefore the graph of Γ^{-1} is the graph of a $1/(1-t)$ -Lipschitz map $\mathbf{s}_t$ pushing μ_t to μ . The uniqueness of this map, even at the level of plans, is proved in Lemma 7.2.1 of [9].

(ii) If $A \in \mathcal{B}(\mathbb{R}^d)$ is $\mathcal{L}^d$ -negligible, then $\mathbf{s}_t(A)$ is also $\mathcal{L}^d$ -negligible, hence μ -negligible. The identity $\mathbf{s}_t \circ \mathbf{t}_t = \mathbf{i}$ μ -a.e. then gives

$$\mu_t(A) = \mu(\mathbf{t}_t^{-1}(A)) \leq \mu(\mathbf{s}_t(A)) = 0.$$

This proves that $\mu_t \ll \mathcal{L}^d$. □

2.5. The continuity equation with locally Lipschitz velocity fields

In this section we collect some results on the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \quad \text{in } \mathbb{R}^d \times (0, T), \quad (2.14)$$

which we will need in the sequel. Here μ_t is a Borel family of probability measures on $\mathbb{R}^d$ defined for t in the open interval $I := (0, T)$, $\mathbf{v} : (x, t) \mapsto \mathbf{v}_t(x) \in \mathbb{R}^d$ is a Borel velocity field such that

$$\int_0^T \int_{\mathbb{R}^d} |\mathbf{v}_t(x)| d\mu_t(x) dt < +\infty, \quad (2.15)$$

and we suppose that (2.14) holds in the sense of distributions, i.e.,

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} (\partial_t \varphi(x, t) + \langle \mathbf{v}_t(x), \nabla_x \varphi(x, t) \rangle) d\mu_t(x) dt &= 0, \\ \forall \varphi \in C_c^\infty(\mathbb{R}^d \times (0, T)). \end{aligned} \quad (2.16)$$

REMARK 2.5 (More general test functions). By a simple regularization argument via convolution, it is easy to show that (2.16) holds if $\varphi \in C_c^1(\mathbb{R}^d \times (0, T))$ as well. Moreover, under condition (2.15), we can also consider bounded test functions φ , with bounded gradient, whose support has a compact projection in $(0, T)$ (that is, the support in x need not be compact): it suffices to approximate φ by $\varphi \chi_R$, where $\chi_R \in C_c^\infty(\mathbb{R}^d)$, $0 \leq \chi_R \leq 1$, $|\nabla \chi_R| \leq 2$ and $\chi_R = 1$ on $B_R(0)$.

First of all we recall some technical preliminaries.

LEMMA 2.6 (Continuous representative). *Let μ_t be a Borel family of probability measures satisfying (2.16) for a Borel vector field $\mathbf{v}_t$ satisfying (2.15). Then there exists a narrowly continuous curve $t \in [0, T] \mapsto \tilde{\mu}_t \in \mathcal{P}(\mathbb{R}^d)$ such that $\mu_t = \tilde{\mu}_t$ for $\mathcal{L}^1$ -a.e. $t \in (0, T)$. Moreover, if $\varphi \in C_c^1(\mathbb{R}^d \times [0, T])$ and $t_1 \leq t_2 \in [0, T]$, we have*

$$\begin{aligned} & \int_{\mathbb{R}^d} \varphi(x, t_2) d\tilde{\mu}_{t_2}(x) - \int_{\mathbb{R}^d} \varphi(x, t_1) d\tilde{\mu}_{t_1}(x) \\ &= \int_{t_1}^{t_2} \int_{\mathbb{R}^d} (\partial_t \varphi + \langle \nabla \varphi, \mathbf{v}_t \rangle) d\mu_t(x) dt. \end{aligned} \quad (2.17)$$

PROOF. Let us take $\varphi(x, t) = \eta(t)\zeta(x)$, $\eta \in C_c^\infty(0, T)$ and $\zeta \in C_c^\infty(\mathbb{R}^d)$; we have

$$- \int_0^T \eta'(t) \left(\int_{\mathbb{R}^d} \zeta(x) d\mu_t(x) \right) dt = \int_0^T \eta(t) \left(\int_{\mathbb{R}^d} \langle \nabla \zeta(x), \mathbf{v}_t(x) \rangle d\mu_t(x) \right) dt,$$

so that the map

$$t \mapsto \mu_t(\zeta) = \int_{\mathbb{R}^d} \zeta(x) d\mu_t(x)$$

belongs to $W^{1,1}(0, T)$ with distributional derivative

$$\dot{\mu}_t(\zeta) = \int_{\mathbb{R}^d} \langle \nabla \zeta(x), \mathbf{v}_t(x) \rangle d\mu_t(x) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, T) \quad (2.18)$$

with

$$|\dot{\mu}_t(\zeta)| \leq V(t) \sup_{\mathbb{R}^d} |\nabla \zeta|, \quad V(t) := \int_{\mathbb{R}^d} |\mathbf{v}_t(x)| d\mu_t(x), \quad V \in L^1(0, T). \quad (2.19)$$

If L_ζ is the set of its Lebesgue points, we know that $\mathcal{L}^1((0, T) \setminus L_\zeta) = 0$. Let us now take a countable set Z which is dense in $C_c^1(\mathbb{R}^d)$ with respect the usual C^1 norm $\|\zeta\|_{C^1} = \sup_{\mathbb{R}^d} (|\zeta|, |\nabla \zeta|)$ and let us set $L_Z := \bigcap_{\zeta \in Z} L_\zeta$. The restriction of the curve μ to L_Z provides a uniformly continuous family of bounded functionals on $C_c^1(\mathbb{R}^d)$, since (2.19) shows

$$|\mu_t(\zeta) - \mu_s(\zeta)| \leq \|\zeta\|_{C^1} \int_s^t V(\lambda) d\lambda \quad \forall s, t \in L_Z.$$

Therefore, it can be extended in a unique way to a continuous curve $\{\tilde{\mu}_t\}_{t \in [0, T]}$ in $[C_c^1(\mathbb{R}^d)]'$. If we show that $\{\mu_t\}_{t \in L_Z}$ is also tight, the extension provides a continuous curve in $\mathcal{P}(\mathbb{R}^d)$.

For, let us consider nonnegative, smooth functions $\zeta_k : \mathbb{R}^d \rightarrow [0, 1]$, $k \in \mathbb{N}$, such that

$$\zeta_k(x) = 1 \quad \text{if } |x| \leq k, \quad \zeta_k(x) = 0 \quad \text{if } |x| \geq k+1, \quad |\nabla \zeta_k(x)| \leq 2.$$

It is not restrictive to suppose that $\zeta_k \in Z$. Applying the previous formula (2.18), for $t, s \in L_Z$ we have

$$|\mu_t(\zeta_k) - \mu_s(\zeta_k)| \leq a_k := 2 \int_0^T \int_{k < |x| < k+1} |\mathbf{v}_\lambda(x)| d\mu_\lambda(x) d\lambda,$$

with $\sum_{k=1}^{+\infty} a_k < +\infty$. For a fixed $s \in L_Z$ and $\varepsilon > 0$, being μ_s tight, we can find $k \in \mathbb{N}$ such that $\mu_s(\zeta_k) > 1 - \varepsilon/2$ and $a_k < \varepsilon/2$. It follows that

$$\mu_t(\overline{B_{k+1}(0)}) \geq \mu_t(\zeta_k) \geq 1 - \varepsilon \quad \forall t \in L_Z.$$

Now we show (2.17). Let us choose $\varphi \in C_c^1(\mathbb{R}^d \times [0, T])$ and set $\varphi_\varepsilon(x, t) = \eta_\varepsilon(t)\varphi(x, t)$, where $\eta_\varepsilon \in C_c^\infty(t_1, t_2)$ such that

$$0 \leq \eta_\varepsilon(t) \leq 1, \quad \lim_{\varepsilon \downarrow 0} \eta_\varepsilon(t) = \chi_{(t_1, t_2)}(t) \quad \forall t \in [0, T], \quad \lim_{\varepsilon \downarrow 0} \eta'_\varepsilon = \delta_{t_1} - \delta_{t_2}$$

in the duality with continuous functions in $[0, T]$. We get

$$\begin{aligned} 0 &= \int_0^T \int_{\mathbb{R}^d} (\partial_t(\eta_\varepsilon \varphi) + \langle \nabla_x(\eta_\varepsilon \varphi), \mathbf{v}_t \rangle) d\mu_t(x) dt \\ &= \int_0^T \eta_\varepsilon(t) \int_{\mathbb{R}^d} (\partial_t \varphi(x, t) + \langle \mathbf{v}_t(x), \nabla_x \varphi(x, t) \rangle) d\mu_t(x) dt \\ &\quad + \int_0^T \eta'_\varepsilon(t) \int_{\mathbb{R}^d} \varphi(x, t) d\tilde{\mu}_t(x) dt. \end{aligned}$$

Passing to the limit as ε vanishes and invoking the continuity of $\tilde{\mu}_t$, we get (2.17). $\square$

LEMMA 2.7 (Time rescaling). *Let $\mathbf{t}: s \in [0, T'] \rightarrow \mathbf{t}(s) \in [0, T]$ be a strictly increasing absolutely continuous map with absolutely continuous inverse $\mathbf{s} := \mathbf{t}^{-1}$. Then $(\mu_t, \mathbf{v}_t)$ is a distributional solution of (2.14) if and only if $\hat{\mu} := \mu \circ \mathbf{t}$, $\hat{\mathbf{v}} := \mathbf{t}' \mathbf{v} \circ \mathbf{t}$, is a distributional solution of (2.14) on $(0, T')$.*

PROOF. By an elementary smoothing argument we can assume that $\mathbf{s}$ is continuously differentiable and $\mathbf{s}' > 0$. We choose $\hat{\varphi} \in C_c^1(\mathbb{R}^d \times (0, T'))$ and we set $\varphi(x, t) := \hat{\varphi}(x, \mathbf{s}(t))$; since $\varphi \in C_c^1(\mathbb{R}^d \times (0, T))$ we have

$$\begin{aligned} 0 &= \int_0^T \int_{\mathbb{R}^d} (\mathbf{s}'(t) \partial_s \hat{\varphi}(x, \mathbf{s}(t)) + \langle \nabla_x \hat{\varphi}(x, \mathbf{s}(t)), \mathbf{v}_t(x) \rangle) d\mu_t(x) dt \\ &= \int_0^T \mathbf{s}'(t) \int_{\mathbb{R}^d} \left(\partial_s \hat{\varphi}(x, \mathbf{s}(t)) + \left\langle \nabla_x \hat{\varphi}(x, \mathbf{s}(t)), \frac{\mathbf{v}_t(x)}{\mathbf{s}'(t)} \right\rangle \right) d\mu_t(x) dt \\ &= \int_0^{T'} \int_{\mathbb{R}^d} \left(\partial_s \hat{\varphi}(x, s) + \langle \nabla_x \hat{\varphi}(x, s), \mathbf{t}'(s) \mathbf{v}_{\mathbf{t}(s)}(x) \rangle \right) d\hat{\mu}_s(x) ds. \end{aligned} \quad \square$$

When the velocity field $\mathbf{v}_t$ is more regular, the classical method of characteristics provides an explicit solution of (2.14). First we recall an elementary result of the theory of ordinary differential equations.

LEMMA 2.8 (The characteristic system of ODE). *Let $\mathbf{v}_t$ be a Borel vector field such that for every compact set $B \subset \mathbb{R}^d$*

$$\int_0^T \left(\sup_B |\mathbf{v}_t| + \text{Lip}(\mathbf{v}_t, B) \right) dt < +\infty. \quad (2.20)$$

Then, for every $x \in \mathbb{R}^d$ and $s \in [0, T]$, the ODE

$$X_s(x, s) = x, \quad \frac{d}{ds} X_t(x, s) = \mathbf{v}_t(X_t(x, s)), \quad (2.21)$$

admits a unique maximal solution defined in an interval $I(x, s)$ relatively open in $[0, T]$ and containing s as (relatively) internal point.

Furthermore, if $t \mapsto |X_t(x, s)|$ is bounded in the interior of $I(x, s)$ then $I(x, s) = [0, T]$; finally, if $\mathbf{v}$ satisfies the global bounds analogous to (2.20)

$$S := \int_0^T \left(\sup_{\mathbb{R}^d} |\mathbf{v}_t| + \text{Lip}(\mathbf{v}_t, \mathbb{R}^d) \right) dt < +\infty, \quad (2.22)$$

then the flow map X satisfies

$$\int_0^T \sup_{x \in \mathbb{R}^d} |\partial_t X_t(x, s)| dt \leq S, \quad \sup_{t, s \in [0, T]} \text{Lip}(X_t(\cdot, s), \mathbb{R}^d) \leq e^S. \quad (2.23)$$

For simplicity, we set $X_t(x) := X_t(x, 0)$ in the particular case $s = 0$ and we denote by $\tau(x) := \sup I(x, 0)$ the length of the maximal time domain of the characteristics leaving from x at $t = 0$.

REMARK 2.9 (The characteristics method for first-order linear PDEs). Characteristics provide a useful representation formula for classical solutions of the backward equation (formally adjoint to (2.14))

$$\partial_t \varphi + \langle \mathbf{v}_t, \nabla \varphi \rangle = \psi \quad \text{in } \mathbb{R}^d \times (0, T); \quad \varphi(x, T) = \varphi_T(x), x \in \mathbb{R}^d, \quad (2.24)$$

when, e.g., $\psi \in C_b^1(\mathbb{R}^d \times (0, T))$, $\varphi_T \in C_b^1(\mathbb{R}^d)$ and $\mathbf{v}$ satisfies the global bounds (2.22), so that maximal solutions are always defined in $[0, T]$. A direct calculation shows that

$$\varphi(x, t) := \varphi_T(X_T(x, t)) - \int_t^T \psi(X_s(x, t), s) ds \quad (2.25)$$

solve (2.24). For $X_s(X_t(x, 0), t) = X_s(x, 0)$ yields

$$\varphi(X_t(x, 0), t) = \varphi_T(X_T(x, 0)) - \int_t^T \psi(X_s(x, 0), s) ds,$$

and differentiating both sides with respect to t we obtain

$$\left[\frac{\partial \varphi}{\partial t} + \langle \mathbf{v}_t, \nabla \varphi \rangle \right] (X_t(x, 0), t) = \psi(X_t(x, 0), t).$$

Since x (and then $X_t(x, 0)$) is arbitrary we conclude that (2.31) is fulfilled.

Now we use characteristics to prove the existence, the uniqueness, and a representation formula of the solution of the continuity equation, under suitable assumption on v .

LEMMA 2.10. *Let $\mathbf{v}_t$ be a Borel velocity field satisfying (2.20), (2.15), let $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$, and let X_t be the maximal solution of the ODE (2.21) (corresponding to $s = 0$). Suppose that for some $\bar{t} \in (0, T]$*

$$\tau(x) > \bar{t} \quad \text{for } \mu_0\text{-a.e. } x \in \mathbb{R}^d. \quad (2.26)$$

Then $t \mapsto \mu_t := (X_t)_\# \mu_0$ is a continuous solution of (2.14) in $[0, \bar{t}]$.

PROOF. The continuity of μ_t follows easily since $\lim_{s \rightarrow t} X_s(x) = X_t(x)$ for μ_0 -a.e. $x \in \mathbb{R}^d$: thus for every continuous and bounded function $\zeta : \mathbb{R}^d \rightarrow \mathbb{R}$ the dominated convergence theorem yields

$$\lim_{s \rightarrow t} \int_{\mathbb{R}^d} \zeta d\mu_s = \lim_{s \rightarrow t} \int_{\mathbb{R}^d} \zeta(X_s(x)) d\mu_0(x) = \int_{\mathbb{R}^d} \zeta(X_t(x)) d\mu_0(x) = \int_{\mathbb{R}^d} \zeta d\mu_t.$$

For any $\varphi \in C_c^\infty(\mathbb{R}^d \times (0, \bar{t}))$ and for μ_0 -a.e. $x \in \mathbb{R}^d$ the maps $t \mapsto \varphi_t(x) := \varphi(X_t(x), t)$ are absolutely continuous in $(0, \bar{t})$, with

$$\dot{\varphi}_t(x) = \partial_t \varphi(X_t(x), t) + \langle \nabla \varphi(X_t(x), t), \mathbf{v}_t(X_t(x)) \rangle = \Lambda(\cdot, t) \circ X_t,$$

where $\Lambda(x, t) := \partial_t \varphi(x, t) + \langle \nabla \varphi(x, t), v_t(x) \rangle$. We thus have

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} |\dot{\varphi}_t(x)| d\mu_0(x) dt &= \int_0^T \int_{\mathbb{R}^d} |\Lambda(X_t(x), t)| d\mu_0(x) dt \\ &= \int_0^T \int_{\mathbb{R}^d} |\Lambda(x, t)| d\mu_t(x) dt \\ &\leq \text{Lip}(\varphi) \left(T + \int_0^T \int_{\mathbb{R}^d} |\mathbf{v}_t(x)| d\mu_t(x) dt \right) \\ &< +\infty \end{aligned}$$

and therefore

$$\begin{aligned}
0 &= \int_{\mathbb{R}^d} \varphi(x, \bar{t}) \, d\mu_{\bar{t}}(x) - \int_{\mathbb{R}^d} \varphi(x, 0) \, d\mu_0(x) \\
&= \int_{\mathbb{R}^d} (\varphi(X_{\bar{t}}(x), \bar{t}) - \varphi(x, 0)) \, d\mu_0(x) \\
&= \int_{\mathbb{R}^d} \left(\int_0^{\bar{t}} \dot{\varphi}_t(x) \, dt \right) d\mu_0(x) \\
&= \int_0^{\bar{t}} \int_{\mathbb{R}^d} (\partial_t \varphi + \langle \nabla \varphi, \mathbf{v}_t \rangle) \, d\mu_t \, dt,
\end{aligned}$$

by a simple application of Fubini's theorem. $\square$

We want to prove that, under reasonable assumptions, in fact *any* solution of (2.14) can be represented as in Lemma 2.10. The first step is a uniqueness theorem for the continuity equation under minimal regularity assumptions on the velocity field. Notice that the only global information on $\mathbf{v}_t$ is (2.27). The proof is based on a classical duality argument (see, for instance, [7,19,41]).

PROPOSITION 2.11 (Uniqueness and comparison for the continuity equation). *Let σ_t be a narrowly continuous family of signed measures solving*

$$\partial_t \sigma_t + \nabla \cdot (\mathbf{v}_t \sigma_t) = 0 \quad \text{in } \mathbb{R}^d \times (0, T),$$

with $\sigma_0 \leq 0$,

$$\int_0^T \int_{\mathbb{R}^d} |\mathbf{v}_t| \, d|\sigma_t| \, dt < +\infty, \tag{2.27}$$

and

$$\int_0^T \left(|\sigma_t|(B) + \sup_B |\mathbf{v}_t| + \text{Lip}(\mathbf{v}_t, B) \right) dt < +\infty$$

for any bounded closed set $B \subset \mathbb{R}^d$. Then $\sigma_t \leq 0$ for any $t \in [0, T]$.

PROOF. Fix $\psi \in C_c^\infty(\mathbb{R}^d \times (0, T))$ with $0 \leq \psi \leq 1$, $R > 0$, and a smooth cut-off function

$$\begin{aligned}
\chi_R(\cdot) &= \chi\left(\frac{\cdot}{R}\right) \in C_c^\infty(\mathbb{R}^d) \quad \text{such that } 0 \leq \chi_R \leq 1, |\nabla \chi_R| \leq \frac{2}{R}, \\
\chi_R &\equiv 1 \quad \text{on } B_R(0) \quad \text{and} \quad \chi_R \equiv 0 \quad \text{on } \mathbb{R}^d \setminus B_{2R}(0).
\end{aligned} \tag{2.28}$$

We define $\mathbf{w}_t$ so that $\mathbf{w}_t = \mathbf{v}_t$ on $B_{2R}(0) \times [0, T]$, $\mathbf{w}_t = 0$ if $t \notin [0, T]$ and

$$\sup_{\mathbb{R}^d} |\mathbf{w}_t| + \text{Lip}(\mathbf{w}_t, \mathbb{R}^d) \leq \sup_{B_{2R}(0)} |\mathbf{v}_t| + \text{Lip}(\mathbf{v}_t, B_{2R}(0)) \quad \forall t \in [0, T]. \quad (2.29)$$

Let $\mathbf{w}_t^\varepsilon$ be obtained from $\mathbf{w}_t$ by a double mollification with respect to the space and time variables: notice that $\mathbf{w}_t^\varepsilon$ satisfy

$$\sup_{\varepsilon \in (0,1)} \int_0^T \left(\sup_{\mathbb{R}^d} |\mathbf{w}_t^\varepsilon| + \text{Lip}(\mathbf{w}_t^\varepsilon, \mathbb{R}^d) \right) dt < +\infty. \quad (2.30)$$

We now build, by the method of characteristics described in Remark 2.9, a smooth solution $\varphi^\varepsilon : \mathbb{R}^d \times [0, T] \rightarrow \mathbb{R}$ of the PDE

$$\frac{\partial \varphi^\varepsilon}{\partial t} + \langle \mathbf{w}_t^\varepsilon, \nabla \varphi^\varepsilon \rangle = \psi \quad \text{in } \mathbb{R}^d \times (0, T), \quad \varphi^\varepsilon(x, T) = 0, x \in \mathbb{R}^d. \quad (2.31)$$

Combining the representation formula (2.25), the uniform bound (2.30), and the estimate (2.23), it is easy to check that $0 \geq \varphi^\varepsilon \geq -T$ and $|\nabla \varphi^\varepsilon|$ is uniformly bounded with respect to ε , t and x .

We insert now the test function $\varphi^\varepsilon \chi_R$ in the continuity equation and take into account that $\sigma_0 \leq 0$ and $\varphi^\varepsilon \leq 0$ to obtain

$$\begin{aligned} 0 &\geq - \int_{\mathbb{R}^d} \varphi^\varepsilon \chi_R d\sigma_0 \\ &= \int_0^T \int_{\mathbb{R}^d} \chi_R \frac{\partial \varphi^\varepsilon}{\partial t} + \langle \mathbf{v}_t, \chi_R \nabla \varphi^\varepsilon + \varphi^\varepsilon \nabla \chi_R \rangle d\sigma_t dt \\ &= \int_0^T \int_{\mathbb{R}^d} \chi_R (\psi + \langle \mathbf{v}_t - \mathbf{w}_t^\varepsilon, \nabla \varphi^\varepsilon \rangle) d\sigma_t dt + \int_0^T \int_{\mathbb{R}^d} \varphi^\varepsilon \langle \nabla \chi_R, \mathbf{v}_t \rangle d\sigma_t dt \\ &\geq \int_0^T \int_{\mathbb{R}^d} \chi_R (\psi + \langle \mathbf{v}_t - \mathbf{w}_t^\varepsilon, \nabla \varphi^\varepsilon \rangle) d\sigma_t dt - \int_0^T \int_{\mathbb{R}^d} |\nabla \chi_R| |\mathbf{v}_t| d|\sigma_t| dt. \end{aligned}$$

Letting $\varepsilon \downarrow 0$ and using the uniform bound on $|\nabla \varphi^\varepsilon|$ and the fact that $\mathbf{w}_t = \mathbf{v}_t$ on $\text{supp } \chi_R \times [0, T]$, we get

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} \chi_R \psi d\sigma_t dt &\leq \int_0^T \int_{\mathbb{R}^d} |\nabla \chi_R| |\mathbf{v}_t| d|\sigma_t| dt \\ &\leq \frac{2}{R} \int_0^T \int_{R \leq |x| \leq 2R} |\mathbf{v}_t| d|\sigma_t| dt. \end{aligned}$$

Eventually letting $R \rightarrow \infty$ we obtain that $\int_0^T \int_{\mathbb{R}^d} \psi d\sigma_t dt \leq 0$. Since ψ is arbitrary the proof is achieved. $\square$

PROPOSITION 2.12 (Representation formula for the continuity equation). *Let μ_t , $t \in [0, T]$, be a narrowly continuous family of Borel probability measures solving the continuity equation (2.14) w.r.t. a Borel vector field $\mathbf{v}_t$ satisfying (2.20) and (2.15). Then for μ_0 -a.e. $x \in \mathbb{R}^d$ the characteristic system (2.21) admits a globally defined solution $X_t(x)$ in $[0, T]$ and*

$$\mu_t = (X_t)_\# \mu_0 \quad \forall t \in [0, T]. \quad (2.32)$$

Moreover, if

$$\int_0^T \int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 d\mu_t(x) dt < +\infty \quad (2.33)$$

then the velocity field $\mathbf{v}_t$ is the time derivative of X_t in the L^2 -sense

$$\lim_{h \downarrow 0} \int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} - \mathbf{v}_t(X_t(x)) \right|^2 d\mu_0(x) dt = 0, \quad (2.34)$$

$$\lim_{h \rightarrow 0} \frac{X_{t+h}(x, t) - x}{h} = \mathbf{v}_t(x) \quad \text{in } L^2(\mu_t; \mathbb{R}^d) \text{ for } \mathcal{L}^1\text{-a.e. } t \in (0, T). \quad (2.35)$$

PROOF. Let $E_s = \{\tau > s\}$ and let us use the fact, proved in Lemma 2.10, that $t \mapsto X_t(\chi_{E_s} \mu_0)$ is a solution of (2.14) in $[0, s]$. By Proposition 2.11 we get also

$$X_t(\chi_{E_s} \mu_0) \leq \mu_t \quad \text{whenever } 0 \leq t \leq s.$$

Using the previous inequality with $s = t$ we can estimate:

$$\begin{aligned} \int_{\mathbb{R}^d} \sup_{(0, \tau(x))} |X_t(x) - x| d\mu_0(x) &\leq \int_{\mathbb{R}^d} \int_0^{\tau(x)} |\dot{X}_t(x)| d\mu_0(x) \\ &= \int_{\mathbb{R}^d} \int_0^{\tau(x)} |\mathbf{v}_t(X_t(x))| d\mu_0(x) \\ &= \int_0^T \int_{E_t} |\mathbf{v}_t(X_t(x))| d\mu_0(x) dt \\ &\leq \int_0^T \int_{\mathbb{R}^d} |\mathbf{v}_t| d\mu_t dt. \end{aligned}$$

It follows that $X_t(x)$ is bounded on $(0, \tau(x))$ for μ_0 -a.e. $x \in \mathbb{R}^d$ and therefore X_t is globally defined in $[0, T]$ for μ_0 -a.e. in $\mathbb{R}^d$. Applying Lemma 2.10 and Proposition 2.11 we obtain (2.32).

Now we observe that the differential quotient $D_h(x, t) := h^{-1}(X_{t+h}(x) - X_t(x))$ can be bounded in $L^2(\mu_0 \times \mathcal{L}^1)$ by

$$\begin{aligned} & \int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} \right|^2 d\mu_0(x) dt \\ &= \int_0^{T-h} \int_{\mathbb{R}^d} \left| \frac{1}{h} \int_0^h \mathbf{v}_{t+s}(X_{t+s}(x)) ds \right|^2 d\mu_0(x) dt \\ &\leq \int_0^{T-h} \int_{\mathbb{R}^d} \frac{1}{h} \int_0^h |\mathbf{v}_{t+s}(X_{t+s}(x))|^2 ds d\mu_0(x) dt \\ &\leq \int_0^T \int_{\mathbb{R}^d} |\mathbf{v}_t(X_t(x))|^2 d\mu_0(x) dt < +\infty. \end{aligned}$$

Since we already know that D_h is pointwise converging to $\mathbf{v}_t \circ X_t$ $\mu_0 \times \mathcal{L}^1$ -a.e. in $\mathbb{R}^d \times (0, T)$, we obtain the strong convergence in $L^2(\mu_0 \times \mathcal{L}^1)$, i.e., (2.34).

Finally, we can consider $t \mapsto X_t(\cdot)$ and $t \mapsto \mathbf{v}_t(X_t(\cdot))$ as maps from $(0, T)$ to $L^2(\mu_0; \mathbb{R}^d)$; (2.34) is then equivalent to

$$\lim_{h \downarrow 0} \int_0^{T-h} \left\| \frac{X_{t+h} - X_t}{h} - \mathbf{v}_t(X_t) \right\|_{L^2(\mu_0; \mathbb{R}^d)}^2 dt = 0,$$

and it shows that $t \mapsto X_t(\cdot)$ belongs to $AC^2(0, T; L^2(\mu_0; \mathbb{R}^d))$. General results for absolutely continuous maps with values in Hilbert spaces yield that X_t is differentiable $\mathcal{L}^1$ -a.e. in $(0, T)$, so that

$$\lim_{h \rightarrow 0} \int_{\mathbb{R}^d} \left| \frac{X_{t+h}(x) - X_t(x)}{h} - \mathbf{v}_t(X_t(x)) \right|^2 d\mu_0(x) = 0 \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, T).$$

Since $X_{t+h}(x) = X_h(X_t(x), t)$, we obtain (2.35). $\square$

Now we state an approximation result for general solution of (2.14) with more regular ones, satisfying the conditions of the previous Proposition 2.12.

LEMMA 2.13 (Approximation by regular curves). *Let μ_t be a time-continuous solution of (2.14) w.r.t. a velocity field satisfying the integrability condition*

$$\int_0^T \int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 d\mu_t(x) dt < +\infty. \quad (2.36)$$

Let $(\rho_\varepsilon) \subset C^\infty(\mathbb{R}^d)$ be a family of strictly positive mollifiers in the x variable (e.g., $\rho_\varepsilon(x) = (2\pi\varepsilon)^{-d/2} \exp(-|x|^2/2\varepsilon)$), and set

$$\mu_t^\varepsilon := \mu_t * \rho_\varepsilon, \quad E_t^\varepsilon := (\mathbf{v}_t \mu_t) * \rho_\varepsilon, \quad \mathbf{v}_t^\varepsilon := \frac{E_t^\varepsilon}{\mu_t^\varepsilon}. \quad (2.37)$$

Then μ_t^ε is a continuous solution of (2.14) w.r.t. $\mathbf{v}_t^\varepsilon$, which satisfies the local regularity assumptions (2.20) and the uniform integrability bounds

$$\int_{\mathbb{R}^d} |\mathbf{v}_t^\varepsilon(x)|^2 d\mu_t^\varepsilon(x) \leq \int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 d\mu_t(x) \quad \forall t \in (0, T). \quad (2.38)$$

Moreover, $E_t^\varepsilon \rightarrow \mathbf{v}_t \mu_t$ narrowly and

$$\lim_{\varepsilon \downarrow 0} \|\mathbf{v}_t^\varepsilon\|_{L^2(\mu_t^\varepsilon; \mathbb{R}^d)} = \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)} \quad \forall t \in (0, T). \quad (2.39)$$

PROOF. With a slight abuse of notation, we are denoting the measure μ_t^ε and its density w.r.t. $\mathcal{L}^d$ by the same symbol. Notice first that $|E^\varepsilon|(t, \cdot)$ and its spatial gradient are uniformly bounded in space by the product of $\|\mathbf{v}_t\|_{L^1(\mu_t)}$ with a constant depending on ε , and the first quantity is integrable in time. Analogously, $|\mu_t^\varepsilon|(t, \cdot)$ and its spatial gradient are uniformly bounded in space by a constant depending on ε . Therefore, as $\mathbf{v}_t^\varepsilon = E_t^\varepsilon / \mu_t^\varepsilon$, the local regularity assumptions (2.20) is fulfilled if

$$\inf_{|x| \leq R, t \in [0, T]} \mu_t^\varepsilon(x) > 0 \quad \text{for any } \varepsilon > 0, R > 0.$$

This property is immediate, since μ_t^ε are continuous w.r.t. t and equi-continuous w.r.t. x , and therefore continuous in both variables.

Lemma 2.14 shows that (2.38) holds. Notice also that μ_t^ε solve the continuity equation

$$\partial_t \mu_t^\varepsilon + \nabla \cdot (\mathbf{v}_t^\varepsilon \mu_t^\varepsilon) = 0 \quad \text{in } \mathbb{R}^d \times (0, T), \quad (2.40)$$

because, by construction, $\nabla \cdot (\mathbf{v}_t^\varepsilon \mu_t^\varepsilon) = \nabla \cdot ((\mathbf{v}_t \mu_t) * \rho_\varepsilon) = (\nabla \cdot (\mathbf{v}_t \mu_t)) * \rho_\varepsilon$. Finally, general lower semicontinuity results on integral functionals defined on measures of the form

$$(E, \mu) \mapsto \int_{\mathbb{R}^d} \left| \frac{E}{\mu} \right|^2 d\mu$$

(see, for instance, Theorem 2.34 and Example 2.36 in [8]) provide (2.39). $\square$

LEMMA 2.14. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and let E be an $\mathbb{R}^m$ -valued measure in $\mathbb{R}^d$ with finite total variation and absolutely continuous with respect to μ . Then*

$$\int_{\mathbb{R}^d} \left| \frac{E * \rho}{\mu * \rho} \right|^2 \mu * \rho dx \leq \int_{\mathbb{R}^d} \left| \frac{E}{\mu} \right|^2 d\mu$$

for any convolution kernel ρ .

PROOF. We use Jensen inequality in the following form: if $\Phi : \mathbb{R}^{m+1} \rightarrow [0, +\infty]$ is convex, l.s.c. and positively 1-homogeneous, then

$$\Phi \left(\int_{\mathbb{R}^d} \psi(x) d\theta(x) \right) \leq \int_{\mathbb{R}^d} \Phi(\psi(x)) d\theta(x)$$

for any Borel map $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^{m+1}$ and any positive and finite measure θ in $\mathbb{R}^d$ (by rescaling θ to be a probability measure and looking at the image measure $\psi_{\#}\theta$ the formula reduces to the standard Jensen inequality). Fix $x \in \mathbb{R}^d$ and apply the inequality above with $\psi := (E/\mu, 1)$, $\theta := \rho(x - \cdot)\mu$ and

$$\Phi(z, t) := \begin{cases} \frac{|z|^2}{t} & \text{if } t > 0, \\ 0 & \text{if } (z, t) = (0, 0), \\ +\infty & \text{if either } t < 0 \text{ or } t = 0, z \neq 0, \end{cases}$$

to obtain

$$\begin{aligned} \left| \frac{E * \rho(x)}{\mu * \rho(x)} \right|^2 \mu * \rho(x) &= \Phi \left(\int_{\mathbb{R}^d} \frac{E}{\mu}(y) \rho(x - y) \, d\mu(y), \int \rho(x - y) \, d\mu(y) \right) \\ &\leq \int_{\mathbb{R}^d} \Phi \left(\frac{E}{\mu}(y), 1 \right) \rho(x - y) \, d\mu(y) \\ &= \int_{\mathbb{R}^d} \left| \frac{E}{\mu} \right|^2(y) \rho(x - y) \, d\mu(y). \end{aligned}$$

An integration with respect to x leads to the desired inequality. $\square$

2.6. The tangent bundle to the Wasserstein space

In this section we endow $\mathcal{P}_2(\mathbb{R}^d)$ with a kind of differential structure, consistent with the metric structure introduced in Section 2.2. Our starting point is the analysis of absolutely continuous curves $\mu_t : (a, b) \rightarrow \mathcal{P}_2(\mathbb{R}^d)$: recall that this concept depends only on the metric structure of $\mathcal{P}_2(\mathbb{R}^d)$, by Definition 2.1. We show in Theorem 2.15 that this class of curves coincides with (distributional) solutions of the continuity equation

$$\frac{\partial}{\partial t} \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \quad \text{in } \mathbb{R}^d \times (a, b).$$

More precisely, given an absolutely continuous curve μ_t , one can find a Borel time-dependent velocity field $\mathbf{v}_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $\|\mathbf{v}_t\|_{L^2(\mu_t)} \leq |\mu'|_t(t)$ for $\mathcal{L}^1$ -a.e. $t \in (a, b)$ and the continuity equation holds. Here $|\mu'|_t(t)$ is the metric derivative of μ_t , defined in (2.2). Conversely, if μ_t solve the continuity equation for some Borel velocity field $\mathbf{w}_t$ with $\int_a^b \|\mathbf{w}_t\|_{L^2(\mu_t)} \, dt < +\infty$, then μ_t is an absolutely continuous curve and $\|\mathbf{w}_t\|_{L^2(\mu_t)} \geq |\mu'|_t(t)$ for $\mathcal{L}^1$ -a.e. $t \in (a, b)$.

As a consequence of Theorem 2.15 we see that among all velocity fields $\mathbf{w}_t$ which produce the same flow μ_t , there is a unique optimal one with smallest $L^2(\mu_t; \mathbb{R}^d)$ -norm, equal to the metric derivative of μ_t ; we view this optimal field as the “tangent” vector field to the curve μ_t . To make this statement more precise, one can show that the minimality of

the L^2 norm of $\mathbf{w}_t$ is characterized by the property

$$\mathbf{w}_t \in \overline{\{\nabla\varphi: \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu_t; \mathbb{R}^d)} \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b). \quad (2.41)$$

The characterization (2.41) of tangent vectors strongly suggests to consider the following tangent bundle to $\mathcal{P}_2(\mathbb{R}^d)$

$$\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) := \overline{\{\nabla\varphi: \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu; \mathbb{R}^d)} \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d), \quad (2.42)$$

endowed with the natural L^2 metric. Moreover, as a consequence of the characterization of absolutely continuous curves in $\mathcal{P}_2(\mathbb{R}^d)$, we recover the Benamou–Brenier (see [15], where the formula was introduced for numerical purposes) formula for the Wasserstein distance:

$$W_2^2(\mu_0, \mu_1) = \min \left\{ \int_0^1 \|\mathbf{w}_t\|_{L^2(\mu_t; \mathbb{R}^d)}^2 dt : \frac{d}{dt} \mu_t + \nabla \cdot (\mathbf{w}_t \mu_t) = 0 \right\}. \quad (2.43)$$

Indeed, for any admissible curve we use the inequality between L^2 norm of $\mathbf{w}_t$ and metric derivative to obtain:

$$\int_0^1 \|\mathbf{w}_t\|_{L^2(\mu_t; \mathbb{R}^d)}^2 dt \geq \int_0^1 |\mu'|^2(t) dt \geq W_2^2(\mu_0, \mu_1).$$

Conversely, since we know that $\mathcal{P}_2(\mathbb{R}^d)$ is a length space, we can use a geodesic μ_t and its tangent vector field $\mathbf{v}_t$ to obtain equality in (2.43). We also show that optimal transport maps belong to $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ under quite general conditions.

In this way we recover in a more general framework the *Riemannian interpretation* of the Wasserstein distance developed by Otto in [74] (see also [57, 73]) and used to study the long time behavior of the porous medium equation. In the original paper [74], (2.43) is derived using formally the concept of Riemannian submersion and the family of maps $\phi \mapsto \phi_\# \mu$ (indexed by $\mu \ll \mathcal{L}^d$) from Arnold's space of diffeomorphisms into the Wasserstein space. In Otto's formalism tangent vectors are rather thought as $s = \frac{d}{dt} \mu_t$ and these vectors are identified, via the continuity equation, with $-D \cdot (\mathbf{v}_s \mu_t)$. Moreover $\mathbf{v}_s$ is chosen to be the gradient of a function ψ_s , so that $D \cdot (\nabla \psi_s \mu_t) = -s$. Then the metric tensor is induced by the identification $s \mapsto \nabla \phi_s$ as follows:

$$\langle s, s' \rangle_{\mu_t} := \int_{\mathbb{R}^d} \langle \nabla \psi_s, \nabla \psi_{s'} \rangle d\mu_t.$$

As noticed in [74], both the identification between tangent vectors and gradients and the scalar product depend on μ_t , and these facts lead to a nontrivial geometry of the Wasserstein space. We prefer instead to consider directly $\mathbf{v}_t$ as the tangent vectors, allowing them to be not necessarily gradients: this leads to (2.42).

Another consequence of the characterization of absolutely continuous curves is a result, given in Proposition 2.20, concerning the infinitesimal behavior of the Wasserstein distance

along absolutely continuous curves μ_t : given the tangent vector field $\mathbf{v}_t$ to the curve, we show that

$$\lim_{h \rightarrow 0} \frac{W_2(\mu_{t+h}, (\mathbf{i} + h\mathbf{v}_t)_\# \mu_t)}{|h|} = 0 \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b).$$

Moreover, the rescaled optimal transport maps between μ_t and μ_{t+h} converge to the transport plan $(\mathbf{i} \times \mathbf{v}_t)_\# \mu_t$ associated to $\mathbf{v}_t$ (see (2.56)). As a consequence, we will obtain in Theorem 2.21 a key formula for the derivative of the map $t \mapsto W_2^2(\mu_t, \nu)$.

THEOREM 2.15 (Absolutely continuous curves in $\mathcal{P}_2(\mathbb{R}^d)$). *Let I be an open interval in $\mathbb{R}$, let $\mu_t : I \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ be an absolutely continuous curve and let $|\mu'| \in L^1(I)$ be its metric derivative, given by (2.2). Then there exists a Borel vector field $\mathbf{v} : (x, t) \mapsto \mathbf{v}_t(x)$ such that*

$$\mathbf{v}_t \in L^2(\mu_t; \mathbb{R}^d), \quad \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)} \leq |\mu'| (t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in I, \quad (2.44)$$

and the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \quad \text{in } \mathbb{R}^d \times I \quad (2.45)$$

holds in the sense of distributions, i.e.,

$$\begin{aligned} \int_I \int_{\mathbb{R}^d} (\partial_t \varphi(x, t) + \langle \mathbf{v}_t(x), \nabla_x \varphi(x, t) \rangle) d\mu_t(x) dt &= 0 \\ \forall \varphi \in C_c^\infty(\mathbb{R}^d \times I). \end{aligned} \quad (2.46)$$

Moreover, for $\mathcal{L}^1$ -a.e. $t \in I$ $\mathbf{v}_t$ belongs to the closure in $L^2(\mu_t, \mathbb{R}^d)$ of the subspace generated by the gradients $\nabla \varphi$ with $\varphi \in C_c^\infty(\mathbb{R}^d)$.

Conversely, if a narrowly continuous curve $\mu_t : I \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ satisfies the continuity equation for some Borel velocity field $\mathbf{w}_t$ with $\|\mathbf{w}_t\|_{L^2(\mu_t; \mathbb{R}^d)} \in L^1(I)$ then $\mu_t : I \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous and $|\mu'| (t) \leq \|\mathbf{w}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$ for $\mathcal{L}^1$ -a.e. $t \in I$.

In particular equality holds in (2.44).

PROOF. Taking into account that any absolutely continuous curve can be reparametrized by arc length (see, for instance, [11]) and Lemma 2.7, we will assume with no loss of generality that $|\mu'| \in L^\infty(I)$ in the proof of the first statement. To fix the ideas, we also assume that $I = (0, 1)$.

First of all we show that for every $\varphi \in C_c^\infty(\mathbb{R}^d)$ the function $t \mapsto \mu_t(\varphi)$ is absolutely continuous, and its derivative can be estimated with the metric derivative of μ_t . Indeed, for $s, t \in I$ and $\mu_{st} \in \Gamma_0(\mu_s, \mu_t)$ we have, using the Hölder inequality,

$$|\mu_t(\varphi) - \mu_s(\varphi)| = \left| \int_{\mathbb{R}^d} (\varphi(y) - \varphi(x)) d\mu_{st} \right| \leq \text{Lip}(\varphi) W_2(\mu_s, \mu_t),$$

whence the absolute continuity follows. In order to estimate more precisely the derivative of $\mu_t(\varphi)$ we introduce the upper semicontinuous and bounded map

$$H(x, y) := \begin{cases} |\nabla \varphi(x)| & \text{if } x = y, \\ \frac{|\varphi(x) - \varphi(y)|}{|x - y|} & \text{if } x \neq y, \end{cases}$$

and notice that, setting $\mu_h = \mu_{(s+h)s}$, we have

$$\begin{aligned} \frac{|\mu_{s+h}(\varphi) - \mu_s(\varphi)|}{|h|} &\leq \frac{1}{|h|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| H(x, y) d\mu_h \\ &\leq \frac{W_2(\mu_{s+h}, \mu_s)}{|h|} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} H^2(x, y) d\mu_h \right)^{1/2}. \end{aligned}$$

If t is a point where $s \mapsto \mu_s$ is metrically differentiable, using the fact that $\mu_h \rightarrow (x, x)_{\#} \mu_t$ narrowly (because their marginals are narrowly converging, any limit point belongs to $\Gamma_0(\mu_t, \mu_t)$ and is concentrated on the diagonal of $\mathbb{R}^d \times \mathbb{R}^d$) we obtain

$$\begin{aligned} \limsup_{h \rightarrow 0} \frac{|\mu_{t+h}(\varphi) - \mu_t(\varphi)|}{|h|} &\leq |\mu'|_t \left(\int_{\mathbb{R}^d} H^2(x, x) d\mu_t \right)^{1/2} \\ &= |\mu'|_t \|\nabla \varphi\|_{L^2(\mu_t; \mathbb{R}^d)}. \end{aligned} \tag{2.47}$$

Set $Q = \mathbb{R}^d \times I$ and let $\mu = \int \mu_t dt \in \mathcal{P}(Q)$ be the measure whose disintegration is $\{\mu_t\}_{t \in I}$. For any $\varphi \in C_c^\infty(Q)$ we have

$$\begin{aligned} &\int_Q \partial_s \varphi(x, s) d\mu(x, s) \\ &= \lim_{h \downarrow 0} \int_Q \frac{\varphi(x, s) - \varphi(x, s - h)}{h} d\mu(x, s) \\ &= \lim_{h \downarrow 0} \int_I \frac{1}{h} \left(\int_{\mathbb{R}^d} \varphi(x, s) d\mu_s(x) - \int_{\mathbb{R}^d} \varphi(x, s) d\mu_{s+h}(x) \right) ds. \end{aligned}$$

Taking into account (2.47), Fatou's lemma yields

$$\begin{aligned} &\left| \int_Q \partial_s \varphi(x, s) d\mu(x, s) \right| \\ &\leq \int_J |\mu'|_s \left(\int_{\mathbb{R}^d} |\nabla \varphi(x, s)|^2 d\mu_s(x) \right)^{1/q} ds \\ &\leq \left(\int_J |\mu'|^2(s) ds \right)^{1/2} \left(\int_Q |\nabla \varphi(x, s)|^2 d\mu(x, s) \right)^{1/2}, \end{aligned} \tag{2.48}$$

where $J \subset I$ is any interval such that $\text{supp } \varphi \subset J \times \mathbb{R}^d$. If $\mathcal{V}$ denotes the closure in $L^2(\mu; \mathbb{R}^d)$ of the subspace $V := \{\nabla \varphi, \varphi \in C_c^\infty(Q)\}$, the previous formula says that the

linear functional $L : V \rightarrow \mathbb{R}$ defined by

$$L(\nabla\varphi) := - \int_Q \partial_s \varphi(x, s) \, d\mu(x, s)$$

can be uniquely extended to a bounded functional on $\mathcal{V}$. Therefore the minimum problem

$$\min \left\{ \frac{1}{2} \int_Q |\mathbf{w}(x, s)|^2 \, d\mu(x, s) - L(\mathbf{w}) : \mathbf{w} \in \mathcal{V} \right\} \quad (2.49)$$

admits a unique solution $\mathbf{v}$ satisfying

$$\int_Q \langle \mathbf{v}(x, s), \nabla\varphi(x, s) \rangle \, d\mu(x, s) = \langle L, \nabla\varphi \rangle \quad \forall \varphi \in C_c^\infty(Q). \quad (2.50)$$

Setting $\mathbf{v}_t(x) = \mathbf{v}(x, t)$ and using the definition of L we obtain (2.46). Moreover, choosing a sequence $(\nabla\varphi_n) \subset V$ converging to $\mathbf{v}$ in $L^2(\mu; \mathbb{R}^d)$, it is easy to show that for $\mathcal{L}^1$ -a.e. $t \in I$ there exists a subsequence $n(i)$ (possibly depending on t) such that $\nabla\varphi_{n(i)}(\cdot, t) \in C_c^\infty(\mathbb{R}^d)$ converge in $L^2(\mu_t; \mathbb{R}^d)$ to $\mathbf{v}(\cdot, t)$.

Finally, choosing an interval $J \subset I$ and $\eta \in C_c^\infty(J)$ with $0 \leq \eta \leq 1$, (2.50) and (2.48) yield

$$\begin{aligned} & \int_Q \eta(s) |\mathbf{v}(x, s)|^2 \, d\mu(x, s) \\ &= \int_Q \eta \langle \mathbf{v}, \mathbf{w} \rangle \, d\mu = \lim_{n \rightarrow \infty} \int_Q \eta \langle \mathbf{v}, \nabla\varphi_n \rangle \, d\mu \\ &= \lim_{n \rightarrow \infty} \langle L, \nabla(\eta\varphi_n) \rangle \leq \| |\mu'| \|_{L^2(J)} \lim_{n \rightarrow \infty} \left(\int_{\mathbb{R}^d \times J} |\nabla\varphi_n|^2 \, d\mu \right)^{1/2} \\ &= \| |\mu'| \|_{L^2(J)} \left(\int_{\mathbb{R}^d \times J} |\mathbf{v}|^2 \, d\mu \right)^{1/2}. \end{aligned}$$

Taking a sequence of smooth approximations of the characteristic function of J we obtain

$$\int_J \int_{\mathbb{R}^d} |\mathbf{v}_s(x)|^2 \, d\mu_s(x) \, ds \leq \int_J |\mu'|^2(s) \, ds, \quad (2.51)$$

and therefore

$$\| \mathbf{v}_t \|_{L^2(\mu_t, \mathbb{R}^d)} \leq |\mu'| (t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in I.$$

Now we show the converse implication. We apply the regularization Lemma 2.13, finding approximations $\mu_t^\varepsilon, \mathbf{w}_t^\varepsilon$ satisfying the continuity equation, the uniform integrability condition (2.15) and the local regularity assumptions (2.20). Therefore, we can apply Proposition 2.12, obtaining the representation formula $\mu_t^\varepsilon = (T_t^\varepsilon)_\# \mu_0^\varepsilon$, where T_t^ε is the maximal solution of the ODE $\dot{T}_t^\varepsilon = \mathbf{w}_t^\varepsilon(T_t^\varepsilon)$ with the initial condition $T_0^\varepsilon = x$ (see Lemma 2.8).

Now, taking into account Lemma 2.14, we estimate

$$\begin{aligned}
\int_{\mathbb{R}^d} |T_{t_2}^\varepsilon(x) - T_{t_1}^\varepsilon(x)|^2 d\mu_0^\varepsilon &\leq (t_2 - t_1) \int_{\mathbb{R}^d} \int_{t_1}^{t_2} |\dot{T}_t^\varepsilon(x)|^2 dt d\mu_0^\varepsilon \\
&= (t_2 - t_1) \int_{t_1}^{t_2} \int_{\mathbb{R}^d} |\mathbf{w}_t^\varepsilon(x)|^2 d\mu_t^\varepsilon dt \\
&\leq (t_2 - t_1) \int_{t_1}^{t_2} \int_{\mathbb{R}^d} |\mathbf{w}_t|^2 d\mu_t dt,
\end{aligned} \tag{2.52}$$

therefore the transport plan $\gamma^\varepsilon := (T_{t_1}^\varepsilon \times T_{t_2}^\varepsilon)_\# \mu_0^\varepsilon \in \Gamma(\mu_{t_1}^\varepsilon, \mu_{t_2}^\varepsilon)$ satisfies

$$W_2^2(\mu_{t_1}^\varepsilon, \mu_{t_2}^\varepsilon) \leq \int_{\mathbb{R}^{2d}} |x - y|^2 d\gamma^\varepsilon \leq (t_2 - t_1) \int_{t_1}^{t_2} \int_{\mathbb{R}^d} |\mathbf{w}_t|^2 d\mu_t dt.$$

Since, for every $t \in I$, μ_t^ε converges narrowly to μ_t as $\varepsilon \rightarrow 0$, a compactness argument (see Lemma 5.2.2 or Proposition 7.1.3 of [9]) gives

$$W_2^2(\mu_{t_1}, \mu_{t_2}) \leq \int_{\mathbb{R}^{2d}} |x - y|^2 d\gamma \leq (t_2 - t_1) \int_{t_1}^{t_2} \int_{\mathbb{R}^d} |\mathbf{w}_t|^2 d\mu_t dt$$

for some optimal transport plan γ between μ_{t_1} and μ_{t_2} . Since t_1 and t_2 are arbitrary this implies that μ_t is absolutely continuous and that its metric derivative is less than $\|\mathbf{w}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$ for $\mathcal{L}^1$ -a.e. $t \in I$. $\square$

Notice that the continuity equation (2.45) involves only the action of $\mathbf{v}_t$ on $\nabla\varphi$ with $\varphi \in C_c^\infty(\mathbb{R}^d)$. Moreover, Theorem 2.15 shows that the minimal norm among all possible velocity fields $\mathbf{w}_t$ is the metric derivative and that $\mathbf{v}_t$ belongs to the L^2 closure of gradients of functions in $C_c^\infty(\mathbb{R}^d)$. These facts suggest a “canonical” choice of $\mathbf{v}_t$ and the following definition of tangent bundle to $\mathcal{P}_2(\mathbb{R}^d)$.

DEFINITION 2.16 (Tangent bundle). Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. We define

$$\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) := \overline{\{\nabla\varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu; \mathbb{R}^d)}.$$

This definition is motivated by the following variational selection principle.

LEMMA 2.17 (Variational selection of the tangent vectors). A vector $\mathbf{v} \in L^2(\mu; \mathbb{R}^d)$ belongs to the tangent space $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ iff

$$\begin{aligned}
\|\mathbf{v} + \mathbf{w}\|_{L^2(\mu; \mathbb{R}^d)} &\geq \|\mathbf{v}\|_{L^2(\mu; \mathbb{R}^d)} \\
\forall \mathbf{w} \in L^2(\mu; \mathbb{R}^d) \text{ such that } \nabla \cdot (\mathbf{w}\mu) &= 0.
\end{aligned} \tag{2.53}$$

In particular, for every $\mathbf{v} \in L^2(\mu; \mathbb{R}^d)$, denoting by $\Pi(\mathbf{v})$ its orthogonal projection on $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$, we have $\nabla \cdot ((\mathbf{v} - \Pi(\mathbf{v}))\mu) = 0$.

PROOF. By the convexity of the L^2 norm, (2.53) holds iff

$$\int_{\mathbb{R}^d} \langle \mathbf{v}, \mathbf{w} \rangle d\mu = 0 \quad \text{for any } \mathbf{w} \in L^2(\mu; \mathbb{R}^d) \text{ such that } \nabla \cdot (\mathbf{w}\mu) = 0. \quad (2.54)$$

As the space of w such that $\nabla \cdot (\mathbf{w}\mu) = 0$ is the orthogonal space to gradients of $C_c^\infty(\mathbb{R}^d)$ functions (in the duality induced by the scalar product of $L^2(\mu; \mathbb{R}^d)$), standard Hilbert duality gives that (2.54) holds iff $\mathbf{v}$ belongs to the L^2 closure of $\{\nabla\phi: \phi \in C_c^\infty(\mathbb{R}^d)\}$. Therefore (2.53) holds iff $\mathbf{v}$ belongs to $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$. $\square$

The remarks above lead also to the following characterization of divergence-free vector fields (we skip the elementary proof of this statement):

PROPOSITION 2.18. *Let $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$. Then $\nabla \cdot (\mathbf{w}\mu) = 0$ iff*

$$\|\mathbf{v} - \mathbf{w}\|_{L^2(\mu; \mathbb{R}^d)} \geq \|\mathbf{v}\|_{L^2(\mu; \mathbb{R}^d)} \quad \forall \mathbf{v} \in \text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d).$$

Moreover equality holds for some $\mathbf{v}$ iff $\mathbf{w} = 0$.

By the characterization (2.54) of $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ we obtain also

$$\text{Tan}_\mu^\perp \mathcal{P}_2(\mathbb{R}^d) = \{\mathbf{v} \in L^2(\mu, \mathbb{R}^d): \nabla \cdot (\mathbf{v}\mu) = 0\}. \quad (2.55)$$

The following two propositions show that the notion of tangent space is consistent with the metric structure, with the continuity equation, and with optimal transport maps (if any).

PROPOSITION 2.19 (Tangent vector to a.c. curves). *Let $\mu_t: I \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ be an absolutely continuous curve and let $\mathbf{v}_t \in L^2(\mu_t; \mathbb{R}^d)$ be such that (2.45) holds. Then $\mathbf{v}_t$ satisfies (2.44) as well if and only if $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ for $\mathcal{L}^1$ -a.e. $t \in I$. The vector $\mathbf{v}_t$ is uniquely determined $\mathcal{L}^1$ -a.e. in I by (2.44) and (2.45).*

PROOF. The uniqueness of $\mathbf{v}_t$ is a straightforward consequence of the linearity with respect to the velocity field of the continuity equation and of the strict convexity of the L^2 norm.

In the proof of Theorem 2.15 we built vector fields $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ satisfying (2.44) and (2.45). By uniqueness, it follows that conditions (2.44) and (2.45) imply $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ for $\mathcal{L}^1$ -a.e. t . $\square$

In the following proposition we recover the tangent vector field to a curve $(\mu_t) \subset \mathcal{P}_2^a(\mathbb{R}^d)$ through the infinitesimal behavior of optimal transport maps along the curve. See Proposition 8.4.6 of [9] for a more general result in the case of curves $(\mu_t) \subset \mathcal{P}_2(\mathbb{R}^d)$.

PROPOSITION 2.20 (Optimal plans along a.c. curves). *Let $\mu_t : I \rightarrow \mathcal{P}_2^a(\mathbb{R}^d)$ be an absolutely continuous curve and let $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ be characterized by Proposition 2.19. Then, for $\mathcal{L}^1$ -a.e. $t \in I$ the following properties hold:*

$$\lim_{h \rightarrow 0} \frac{1}{h} (\mathbf{t}_{\mu_t}^{\mu_{t+h}} - \mathbf{i}) = \mathbf{v}_t \quad \text{in } L^2(\mu_t; \mathbb{R}^d), \quad (2.56)$$

where $\mathbf{t}_{\mu_t}^{\mu_{t+h}}$ is the unique optimal transport map between μ_t and μ_{t+h} , and

$$\lim_{h \rightarrow 0} \frac{W_2(\mu_{t+h}, (\mathbf{i} + h\mathbf{v}_t)_{\#} \mu_t)}{|h|} = 0. \quad (2.57)$$

PROOF. Let $\mathcal{D} \subset C_c^\infty(\mathbb{R}^d)$ be a countable set with the following property: for any integer $R > 0$ and any $\varphi \in C_c^\infty(\mathbb{R}^d)$ with $\text{supp } \varphi \subset B_R$ there exist $(\varphi_n) \subset \mathcal{D}$ with $\text{supp } \varphi_n \subset B_R$ and $\varphi_n \rightarrow \varphi$ in $C^1(\mathbb{R}^d)$.

We fix $t \in I$ such that $W_2(\mu_{t+h}, \mu_t)/|h| \rightarrow |\mu'| (t) = \|\mathbf{v}_t\|_{L^2(\mu_t)}$ and

$$\lim_{h \rightarrow 0} \frac{\mu_{t+h}(\varphi) - \mu_t(\varphi)}{h} = \int_{\mathbb{R}^d} \langle \nabla \varphi, \mathbf{v}_t \rangle d\mu_t \quad \forall \varphi \in \mathcal{D}. \quad (2.58)$$

Since $\mathcal{D}$ is countable, the metric differentiation theorem implies that both conditions are fulfilled for $\mathcal{L}^1$ -a.e. $t \in I$. Set

$$\mathbf{s}_h := \frac{\mathbf{t}_{\mu_t}^{\mu_{t+h}} - \mathbf{i}}{h}$$

and fix $\varphi \in \mathcal{D}$ and a weak limit point $\mathbf{s}_0$ of $\mathbf{s}_h$ as $h \rightarrow 0$. We use the identity

$$\begin{aligned} \frac{\mu_{t+h}(\varphi) - \mu_t(\varphi)}{h} &= \frac{1}{h} \int_{\mathbb{R}^d} \varphi(\mathbf{t}_{\mu_t}^{\mu_{t+h}}(x)) - \varphi(x) d\mu_t \\ &= \frac{1}{h} \int_{\mathbb{R}^d} \varphi(x + h\mathbf{s}_h(x)) - \varphi(x) d\mu_h \\ &= h \int_{\mathbb{R}^d} \langle \nabla \varphi(x), \mathbf{s}_h(x) \rangle + \omega_x(h) d\mu_h \end{aligned}$$

with $\omega_x(h)$ bounded and infinitesimal as $h \rightarrow 0$, to obtain

$$\int_{\mathbb{R}^d} \langle \nabla \varphi, \mathbf{v}_t \rangle d\mu_t = \int_{\mathbb{R}^d} \langle \nabla \varphi, \mathbf{s}_0 \rangle d\mu_t(x).$$

By the density of $\mathcal{D}$ it follows that

$$\nabla \cdot ((\mathbf{s}_0 - \mathbf{v}_t)\mu_t) = 0. \quad (2.59)$$

We now claim that

$$\int_{\mathbb{R}^d} |\mathbf{s}_0|^2 d\mu_t(x) \leq [|\mu'| (t)]^2. \quad (2.60)$$

Indeed

$$\begin{aligned}
\int_{\mathbb{R}^d} |\mathbf{s}_0|^2 d\mu_t(x) &\leq \liminf_{h \rightarrow 0} \int_{\mathbb{R}^d} |\mathbf{s}_h|^2 d\mu_t \\
&= \liminf_{h \rightarrow 0} \frac{1}{h^2} \int_{\mathbb{R}^d} |\mathbf{t}_{\mu_t}^{\mu_t+h}(x) - x|^2 d\mu_t \\
&= \liminf_{h \rightarrow 0} \frac{W_2^2(\mu_{t+h}, \mu_t)}{h^2} = |\mu'|^2(t).
\end{aligned}$$

From (2.60) we obtain that $\|\mathbf{s}_0\|_{L^2(\mu_t; \mathbb{R}^d)} \leq [|\mu'|](t) = \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$. Therefore Proposition 2.18 entails that $\mathbf{s}_0 = \mathbf{v}_t$. Moreover, the first inequality above is strict if $\mathbf{s}_h$ converge weakly, but not strongly, to $\mathbf{s}_0$. Therefore (2.56) holds.

Now we show (2.57). By (2.8) we can estimate the distance between μ_{t+h} and $(\mathbf{i} + h\mathbf{v}_t)_\# \mu_t$ with $\|\mathbf{i} + h\mathbf{v}_t - \mathbf{t}_{\mu_t}^{\mu_t+h}\|_{L^2(\mu_t; \mathbb{R}^d)}$, and because of (2.56) this norm tends to 0 faster than h . $\square$

As an application of (2.57) we are now able to show the $\mathcal{L}^1$ -a.e. differentiability of $t \mapsto W_2(\mu_t, \sigma)$ along absolutely continuous curves μ_t , with $\mu_t \in \mathcal{P}_2^a(\mathbb{R}^d)$.

THEOREM 2.21 (Generic differentiability of $W_2(\mu_t, \sigma)$). *Let $\mu_t : I \rightarrow \mathcal{P}_2^a(\mathbb{R}^d)$ be an absolutely continuous curve, let $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ and let $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ be its tangent vector field, characterized by Proposition 2.19. Then*

$$\frac{d}{dt} W_2^2(\mu_t, \sigma) = 2 \int_{\mathbb{R}^d} \langle x - \mathbf{t}_{\mu_t}^\sigma(x), \mathbf{v}_t(x) \rangle d\mu_t(x) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in I. \quad (2.61)$$

PROOF. We show that the stated property is true at any $\bar{t}$ where (2.57) holds and the derivative of $t \mapsto W_2(\mu_t, \sigma)$ exists (recall that this map is absolutely continuous). Due to (2.57), we know that the limit

$$L := \lim_{h \rightarrow 0} \frac{W_2^2((\mathbf{i} + h\mathbf{v}_{\bar{t}})_\# \mu_{\bar{t}}, \sigma) - W_2^2(\mu_{\bar{t}}, \sigma)}{h}$$

exists and coincides with $\frac{d}{dt} W_2^2(\mu_t, \sigma)$ evaluated at $t = \bar{t}$, and we have to show that it is equal to the left-hand side in (2.61).

Using the transport maps $\mathbf{i} + h\mathbf{v}_{\bar{t}}, \mathbf{t}_{\mu_{\bar{t}}}^\sigma$ to estimate from above $W_2((\mathbf{i} + h\mathbf{v}_{\bar{t}})_\# \mu_{\bar{t}}, \sigma)$, we get

$$\begin{aligned}
W_2^2((\mathbf{i} + h\mathbf{v}_{\bar{t}})_\# \mu_{\bar{t}}, \sigma) &\leq \int_{\mathbb{R}^d} |(\mathbf{i} + h\mathbf{v}_{\bar{t}}) - \mathbf{t}_{\mu_{\bar{t}}}^\sigma|^2 d\mu_{\bar{t}} \\
&= 2h \int_{\mathbb{R}^d} \langle \mathbf{i} - \mathbf{t}_{\mu_{\bar{t}}}^\sigma, \mathbf{v}_{\bar{t}} \rangle d\mu_{\bar{t}} + o(h) + \int_{\mathbb{R}^d} |\mathbf{i} - \mathbf{t}_{\mu_{\bar{t}}}^\sigma|^2 d\mu_{\bar{t}}.
\end{aligned}$$

Subtracting the last integral, dividing both sides by h and taking limits as $h \downarrow 0$ or $h \uparrow 0$ we obtain

$$L \leq 2 \int_{\mathbb{R}^d} \langle x - \mathbf{t}_{\mu_t}^\sigma(x), \mathbf{v}_t(x) \rangle d\mu_t(x) \leq L. \quad \square$$

The argument in the previous proof leads to the so-called super-differentiability property of the Wasserstein distance, a theme used in many papers on this subject (see in particular [67] and Chapter 10 of [9]). Finally, we compare the tangent space arising from the closure of gradients of smooth compactly supported function with the tangent space built using optimal maps. Proposition 2.20 suggests indeed another possible definition of tangent cone to a measure $\mu \in \mathcal{P}_2^a(\mathbb{R}^d)$: we define

$$\text{Tan}_\mu^r \mathcal{P}_2(\mathbb{R}^d) := \overline{\{\lambda(\mathbf{t}_\mu^v - \mathbf{i}) : v \in \mathcal{P}_2(\mathbb{R}^d), \lambda > 0\}}^{L^2(\mu; \mathbb{R}^d)}. \quad (2.62)$$

As a matter of fact, the two concepts coincide (see also Section 8.5 of [9] for a more general statement).

THEOREM 2.22. *For any $\mu \in \mathcal{P}_2^a(\mathbb{R}^d)$ we have $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) = \text{Tan}_\mu^r \mathcal{P}_2(\mathbb{R}^d)$.*

PROOF. We show first that optimal transport maps $\mathbf{t} = \mathbf{t}_{\mu_t}^\sigma$ belong to $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$. Assume that $\text{supp } \sigma$ is contained in $\bar{B}_R(0)$ for some $R > 0$. We know that we can represent $\mathbf{t} = \nabla \varphi$, where φ is a Lipschitz convex function. We consider now the mollified functions φ_ε . A truncation argument enabling an approximation by gradients with compact support gives that $\nabla \varphi_\varepsilon$ belong to $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$. Due to the absolute continuity of μ it is immediate to check using the dominated convergence theorem that $\nabla \varphi_\varepsilon$ converge to $\nabla \varphi$ in $L^2(\mu; \mathbb{R}^d)$, therefore $\nabla \varphi \in \text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ as well. In the case when the support of σ is not bounded we approximate σ in $\mathcal{P}_2(\mathbb{R}^d)$ by measures with compact support (details are worked out in Lemma 8.5.3 of [9]).

Now we show the opposite inclusion: if $\varphi \in C_c^\infty(\mathbb{R}^d)$ it is always possible to choose $\lambda > 0$ such that $x \mapsto \frac{1}{2}|x|^2 + \lambda^{-1}\varphi(x)$ is convex. Therefore $\mathbf{r} := \mathbf{i} + \lambda^{-1}\nabla \varphi$ is the optimal map between μ and $\nu := \mathbf{r}_\# \mu$; by (2.62) we obtain that $\nabla \varphi = \lambda(\mathbf{r} - \mathbf{i})$ belongs to $\text{Tan}_\mu^r \mathcal{P}_2(\mathbb{R}^d)$. $\square$

3. Convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$

The importance of geodesically convex functionals in Wasserstein spaces was firstly pointed out by McCann [66], who introduced the three basic examples we will discuss in detail in Sections 3.4, 3.6 and 3.8. His original motivation was to prove the uniqueness of the minimizer of an energy functional which results from the sum of the above three contributions.

Applications of this idea have been given to (im)prove many deep functional (Brunn–Minkowski, Gaussian, (logarithmic) Sobolev, isoperimetric, etc.) inequalities: we refer to Villani's book [86], Chapter 6 (see also the survey [49]) for a detailed account on this topic. Connections with evolution equations have also been exploited [2, 29, 70, 74, 75], mainly to study the asymptotic decay of the solution to the equilibrium.

From our point of view, convexity is a crucial tool to study the well posedness and the basic regularity properties of gradient flows. Thus in this section we discuss the basic notions and properties related to this concept: the first part of Section 3.1 is devoted to fixing the notion of convexity along geodesics in $\mathcal{P}_2(\mathbb{R}^d)$.

Section 3.2 discusses in great generality the main examples of geodesically convex functionals: potential, interaction and internal energy. We consider also the convexity properties of the map $\mu \mapsto -W_2^2(\mu, \nu)$ and its geometric implications.

In the last section we give a closer look to the convexity properties of general relative entropy functionals, showing that they are strictly related to the log-concavity of the reference measures.

3.1. λ -geodesically convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$

In McCann's approach, a functional $\phi : \mathcal{P}_2^a(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ is *displacement convex* if

$$\text{setting } \mu_t^{1 \rightarrow 2} := (\mathbf{i} + t(\mathbf{t} - \mathbf{i}))_{\#} \mu^1, \text{ with } \mathbf{t} = \mathbf{t}_{\mu^1}^{\mu^2}, \text{ the map } t \in [0, 1] \mapsto \phi(\mu_t^{1 \rightarrow 2}) \text{ is convex, } \forall \mu^1, \mu^2 \in \mathcal{P}_2^a(\mathbb{R}^d). \quad (3.1)$$

We have seen that the curve $\mu_t^{1 \rightarrow 2}$ is the unique constant speed geodesic connecting μ^1 to μ^2 ; therefore the following definition seems natural, when we consider functionals whose domain contains general probability measures.

DEFINITION 3.1 (λ -convexity along geodesics). Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$. Given $\lambda \in \mathbb{R}$, we say that ϕ is λ -geodesically convex in $\mathcal{P}_2(\mathbb{R}^d)$ if for every couple $\mu^1, \mu^2 \in \mathcal{P}_2(\mathbb{R}^d)$ there exists $\mu \in \Gamma_0(\mu^1, \mu^2)$ such that

$$\phi(\mu_t^{1 \rightarrow 2}) \leq (1-t)\phi(\mu^1) + t\phi(\mu^2) - \frac{\lambda}{2}t(1-t)W_2^2(\mu^1, \mu^2) \quad \forall t \in [0, 1], \quad (3.2)$$

where $\mu_t^{1 \rightarrow 2} = ((1-t)\pi^1 + t\pi^2)_{\#} \mu$, π^1, π^2 being the projections onto the first and the second coordinate in $\mathbb{R}^d \times \mathbb{R}^d$, respectively.

REMARK 3.2 (The map $t \mapsto \phi(\mu_t^{1 \rightarrow 2})$ is λ -convex). The standard definition of λ -convexity for a map $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ requires

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y) - \frac{\lambda}{2}t(1-t)|x - y|^2 \quad \forall t \in [0, 1], x, y \in \mathbb{R}^n \quad (3.3)$$

(equivalently, if φ is continuous, one might ask that $D^2\varphi \geq \lambda I$ in the sense of distributions). The definition of λ -convexity expressed through (3.2) implies that

$$\text{the map } t \in [0, 1] \mapsto \phi(\mu_t^{1 \rightarrow 2}) \text{ is } \lambda W_2^2(\mu^1, \mu^2)\text{-convex,} \quad (3.4)$$

thus recovering an (apparently) stronger and more traditional form. This equivalence follows easily by the fact that for $t_1 < t_2$ in $[0, 1]$ with $\{t_1, t_2\} \neq \{0, 1\}$ the plan $((1 - t_1)\pi^1 + t_1\pi^2) \times ((1 - t_2)\pi^1 + t_2\pi^2))_{\#}\mu$ is the *unique* element of $\Gamma_0(\mu_{t_1}^{1 \rightarrow 2}, \mu_{t_2}^{1 \rightarrow 2})$.

Let us discuss now the convexity properties of the squared Wasserstein distance. In the one-dimensional case it can be easily shown (see Theorem 6.0.2 of [9]) that $\mathcal{P}_2(\mathbb{R})$ is isometrically isomorphic to a closed convex subset of an Hilbert space: precisely the space of nondecreasing functions in $(0, 1)$ (the inverses of distribution functions), viewed as a subset of $L^2(0, 1)$. Thus the Wasserstein distance in $\mathbb{R}$ satisfies the generalized parallelogram rule

$$\begin{aligned} W_2^2(\mu^1, \mu_t^{2 \rightarrow 3}) &= (1 - t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) - t(1 - t)W_2^2(\mu^2, \mu^3) \\ \forall t \in [0, 1], \mu^1, \mu^2, \mu^3 &\in \mathcal{P}_2(\mathbb{R}). \end{aligned} \quad (3.5)$$

On the other hand, if the ambient space has dimension ≥ 2 the following example shows that there is no constant λ such that $W_2^2(\cdot, \mu^1)$ is λ -convex along geodesics.

EXAMPLE 3.3 (The squared distance function is not λ -convex). Let $d = 2$ and

$$\mu^2 := \frac{1}{2}(\delta_{(0,0)} + \delta_{(2,1)}), \quad \mu^3 := \frac{1}{2}(\delta_{(0,0)} + \delta_{(-2,1)}).$$

It is easy to check that the unique optimal map $\mathbf{r}$ pushing μ^2 to μ^3 maps $(0, 0)$ in $(-2, 1)$ and $(2, 1)$ in $(0, 0)$, therefore there is a unique constant speed geodesic joining the two measures, given by

$$\mu_t^{2 \rightarrow 3} := \frac{1}{2}(\delta_{(-2t,t)} + \delta_{(2-2t,1-t)}), \quad t \in [0, 1].$$

Choosing $\mu^1 := \frac{1}{2}(\delta_{(0,0)} + \delta_{(0,-2)})$, there are two maps $\mathbf{r}_t, \mathbf{s}_t$ pushing μ^1 to $\mu_t^{2 \rightarrow 3}$, given by

$$\begin{aligned} \mathbf{r}_t(0, 0) &= (-2t, t), & \mathbf{r}_t(0, -2) &= (2 - 2t, 1 - t), \\ \mathbf{s}_t(0, 0) &= (2 - 2t, 1 - t), & \mathbf{s}_t(0, -2) &= (-2t, t). \end{aligned}$$

Therefore

$$W_2^2(\mu_t^{2 \rightarrow 3}, \mu^1) = \min \left\{ 5t^2 - 7t + \frac{13}{2}, 5t^2 - 3t + \frac{9}{2} \right\}$$

has a concave cusp at $t = 1/2$ and therefore is not λ -convex along the geodesic $\mu_t^{2 \rightarrow 3}$ for any $\lambda \in \mathbb{R}$.

3.2. Examples of convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$

In this section we introduce the main classes of geodesically convex functionals.

EXAMPLE 3.4 (Potential energy). Let $V : \mathbb{R}^d \rightarrow (-\infty, +\infty]$ be a proper, lower semicontinuous function whose negative part has a quadratic growth, i.e.,

$$V(x) \geq -A - B|x|^2 \quad \forall x \in \mathbb{R}^d \text{ for some } A, B \in \mathbb{R}^+. \quad (3.6)$$

In $\mathcal{P}_2(\mathbb{R}^d)$ we define

$$\mathcal{V}(\mu) := \int_{\mathbb{R}^d} V(x) d\mu(x). \quad (3.7)$$

Evaluating $\mathcal{V}$ on Dirac's masses we check that $\mathcal{V}$ is proper; since V^- has at most quadratic growth Lemma 1.2 gives that $\mathcal{V}$ is lower semicontinuous in $\mathcal{P}_2(\mathbb{R}^d)$. If V is bounded from below we have even lower semicontinuity w.r.t. narrow convergence.

The following simple proposition shows that $\mathcal{V}$ is convex along all interpolating curves induced by admissible plans; choosing optimal plans one obtains in particular that $\mathcal{V}$ is convex along geodesics.

PROPOSITION 3.5 (Convexity of $\mathcal{V}$). *If V is λ -convex then for every $\mu^1, \mu^2 \in D(\mathcal{V})$ and $\mu \in \Gamma(\mu^1, \mu^2)$ we have*

$$\begin{aligned} \mathcal{V}(\mu_t^{1 \rightarrow 2}) &\leq (1-t)\mathcal{V}(\mu^1) + t\mathcal{V}(\mu^2) \\ &\quad - \frac{\lambda}{2}t(1-t) \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^2 d\mu(x_1, x_2). \end{aligned} \quad (3.8)$$

In particular $\mathcal{V}$ is λ -convex along geodesics.

PROOF. Since V is bounded from below either by a continuous affine functional (if $\lambda \geq 0$) or by a quadratic function (if $\lambda < 0$) its negative part satisfies (3.6); therefore the definition (3.7) makes sense.

Integrating (3.3) along any admissible transport plan $\mu \in \Gamma(\mu^1, \mu^2)$ with $\mu^1, \mu^2 \in D(\mathcal{V})$ we obtain (3.8), since

$$\begin{aligned} \mathcal{V}(\mu_t^{1 \rightarrow 2}) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} V((1-t)x_1 + tx_2) d\mu(x_1, x_2) \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \left((1-t)V(x_1) + tV(x_2) - \frac{\lambda}{2}t(1-t)|x_1 - x_2|^2 \right) d\mu(x_1, x_2) \\ &= (1-t)\mathcal{V}(\mu^1) + t\mathcal{V}(\mu^2) - \frac{\lambda}{2}t(1-t) \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^2 d\mu(x_1, x_2). \end{aligned}$$

Since $\mathcal{V}(\delta_x) = V(x)$, it is easy to check that the conditions on V are also necessary for the validity of the previous proposition. $\square$

EXAMPLE 3.6 (Interaction energy). Let us fix an integer $k > 1$ and let us consider a lower semicontinuous function $W : \mathbb{R}^{kd} \rightarrow (-\infty, +\infty]$, whose negative part satisfies the usual quadratic growth condition. Denoting by $\mu^{\times k}$ the measure $\mu \times \mu \times \cdots \times \mu$ on $\mathbb{R}^{kd}$, we set

$$\mathcal{W}_k(\mu) := \int_{\mathbb{R}^{kd}} W(x_1, x_2, \dots, x_k) d\mu^{\times k}(x_1, x_2, \dots, x_k). \quad (3.9)$$

If

$$\exists x \in \mathbb{R}^d: \quad W(x, x, \dots, x) < +\infty, \quad (3.10)$$

then $\mathcal{W}_k$ is proper; its lower semicontinuity follows from the fact that

$$\mu_n \rightarrow \mu \quad \text{in } \mathcal{P}_2(\mathbb{R}^d) \implies \mu_n^{\times k} \rightarrow \mu^{\times k} \quad \text{in } \mathcal{P}_2(\mathbb{R}^{kd}). \quad (3.11)$$

Here the typical example is $k = 2$ and $W(x_1, x_2) := \tilde{W}(x_1 - x_2)$ for some $\tilde{W} : \mathbb{R}^d \rightarrow (-\infty, +\infty]$ with $\tilde{W}(0) < +\infty$.

PROPOSITION 3.7 (Convexity of $\mathcal{W}$). *If W is convex then the functional $\mathcal{W}_k$ is convex along the interpolating curve $\mu_t^{1 \rightarrow 2}$ induced by any $\mu \in \Gamma(\mu^1, \mu^2)$, in $\mathcal{P}_2(\mathbb{R}^d)$.*

PROOF. Observe that $\mathcal{W}_k$ is the restriction to the subset

$$\mathcal{P}_2^{\times}(\mathbb{R}^{kd}) := \{\mu^{\times k} : \mu \in \mathcal{P}_2(\mathbb{R}^d)\}$$

of the potential energy functional $\mathcal{W}$ on $\mathcal{P}_2(\mathbb{R}^{kd})$ given by

$$\mathcal{W}(\mu) := \int_{\mathbb{R}^{kd}} W(x_1, \dots, x_k) d\mu(x_1, \dots, x_k).$$

We consider the linear permutation of coordinates $P : (\mathbb{R}^{2d})^k \rightarrow (\mathbb{R}^{kd})^2$ defined by

$$P((x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)) := ((x_1, \dots, x_k), (y_1, \dots, y_k)).$$

If $\mu \in \Gamma(\mu_1, \mu_2)$ then it is easy to check that $P_{\#}\mu^{\times k} \in \Gamma(\mu_1^{\times k}, \mu_2^{\times k}) \subset \mathcal{P}((\mathbb{R}^{kd})^2)$ and

$$(\pi_t^{1 \rightarrow 2})_{\#} P_{\#}(\mu^{\times k}) = P_{\#}((\pi_t^{1 \rightarrow 2})_{\#} \mu)^{\times k}.$$

Therefore all the convexity properties of $\mathcal{W}_k$ follow from the corresponding ones of $\mathcal{W}$. $\square$

EXAMPLE 3.8 (Internal energy). Let $F : [0, +\infty) \rightarrow (-\infty, +\infty]$ be a proper, lower semicontinuous convex function such that

$$F(0) = 0, \quad \liminf_{s \downarrow 0} \frac{F(s)}{s^{\alpha}} > -\infty \quad \text{for some } \alpha > \frac{d}{d+2}. \quad (3.12)$$

We consider the functional $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ defined by

$$\mathcal{F}(\mu) := \begin{cases} \int_{\mathbb{R}^d} F(u(x)) \, d\mathcal{L}^d(x) & \text{if } \mu = u \cdot \mathcal{L}^d \in \mathcal{P}_2^a(\mathbb{R}^d), \\ +\infty & \text{otherwise.} \end{cases} \quad (3.13)$$

REMARK 3.9 (The meaning of condition (3.12)). Condition (3.12) simply guarantees that the negative part of $F(\mu)$ is integrable in $\mathbb{R}^d$. For, let us observe that there exist nonnegative constants c_1, c_2 such that the negative part of F satisfies

$$F^-(s) \leq c_1 s + c_2 s^\alpha \quad \forall s \in [0, +\infty),$$

and it is not restrictive to suppose $\alpha \leq 1$. Since $\mu = u \mathcal{L}^d \in \mathcal{P}_2(\mathbb{R}^d)$ and $2\alpha/(1-\alpha) > d$ we have

$$\begin{aligned} & \int_{\mathbb{R}^d} u^\alpha(x) \, d\mathcal{L}^d(x) \\ &= \int_{\mathbb{R}^d} u^\alpha(x) (1+|x|)^{2\alpha} (1+|x|)^{-2\alpha} \, d\mathcal{L}^d(x) \\ &\leq \left(\int_{\mathbb{R}^d} u(x) (1+|x|)^2 \, d\mathcal{L}^d(x) \right)^\alpha \left(\int_{\mathbb{R}^d} (1+|x|)^{-2\alpha/(1-\alpha)} \, d\mathcal{L}^d(x) \right)^{1-\alpha} \\ &< +\infty \end{aligned}$$

and therefore $F^-(u) \in L^1(\mathbb{R}^d)$.

REMARK 3.10 (Lower semicontinuity of $\mathcal{F}$). General results on integral functionals (see, for instance, [8]) show that $\mathcal{F}$ is narrowly lower semicontinuous if F is nonnegative and has a superlinear growth at infinity. Indeed, under this assumption sequences $\mu_n = u_n \mathcal{L}^d$ on which $\mathcal{F}$ is bounded have the property that (u_n) is sequentially weakly relatively compact in $L^1(\mathbb{R}^d)$, and the convexity of $\mathcal{F}$ together with the lower semicontinuity of F ensure the sequential lower semicontinuity with respect to the weak L^1 topology.

In the next proposition we prove the geodesic convexity of the internal energy functional (3.13) by using the change of variable formula (1.24). This was first shown by McCann [66] with a different argument.

PROPOSITION 3.11 (Convexity of $\mathcal{F}$). *If F has a superlinear growth at infinity and*

$$\text{the map } s \mapsto s^d F(s^{-d}) \text{ is convex and nonincreasing in } (0, +\infty), \quad (3.14)$$

then the functional $\mathcal{F}$ is convex along geodesics in $\mathcal{P}_2(\mathbb{R}^d)$.

PROOF. We consider two measures $\mu^i = u^i \mathcal{L}^d \in D(\mathcal{F})$, $i = 1, 2$, and the optimal transport map $\mathbf{r}$ such that $\mathbf{r}_\# \mu^1 = \mu^2$. Setting $\mathbf{r}_t := (1-t)\mathbf{i} + t\mathbf{r}$, by the characterization of constant

speed geodesics we know that $\mathbf{r}_t$ is the optimal transport map between μ^1 and $\mu_t := \mathbf{r}_t \# \mu^1$ for any $t \in [0, 1]$, and $\mu_t = u_t \# \mathcal{L}^d \in \mathcal{P}_2^a(\mathbb{R}^d)$, with

$$u_t(\mathbf{r}_t(x)) = \frac{u^1(x)}{\det \nabla \mathbf{r}_t(x)} \quad \text{for } \mu^1\text{-a.e. } x \in \mathbb{R}^d.$$

By (1.24) it follows that

$$\mathcal{F}(\mu_t) = \int_{\mathbb{R}^d} F(u_t(y)) \, dy = \int_{\mathbb{R}^d} F\left(\frac{u^1(x)}{\det \nabla \mathbf{r}_t(x)}\right) \det \nabla \mathbf{r}_t(x) \, dx.$$

Since for a diagonalizable map D with nonnegative eigenvalues

$$t \mapsto \det((1-t)I + tD)^{1/d} \quad \text{is concave in } [0, 1], \quad (3.15)$$

the integrand above may be seen as the composition of the convex and nonincreasing map $s \mapsto s^d F(u^1(x)/s^d)$ and of the concave map in (3.15), so that the resulting map is convex in $[0, 1]$ for μ^1 -a.e. $x \in \mathbb{R}^d$. Thus we have

$$F\left(\frac{u^1(x)}{\det \nabla \mathbf{r}_t(x)}\right) \det \nabla \mathbf{r}_t(x) \leq (1-t)F(u^1(x)) + tF(u^2(x))$$

and the thesis follows by integrating this inequality in $\mathbb{R}^d$. $\square$

In order to express (3.14) in a different way, we introduce the function

$$\begin{aligned} L_F(z) &:= zF'(z) - F(z) \\ \text{which satisfies } -L_F(e^{-z})e^z &= \frac{d}{dz} F(e^{-z})e^z; \end{aligned} \quad (3.16)$$

denoting by $\widehat{F}$ the modified function $F(e^{-z})e^z$ we have the simple relation

$$\begin{aligned} \widehat{L}_F(z) &= -\frac{d}{dz} \widehat{F}(z), \quad \widehat{L}_F^2(z) = -\frac{d}{dz} \widehat{L}_F(z) = \frac{d^2}{dz^2} \widehat{F}(z), \\ \text{where } L_F^2(z) &:= L_{L_F}(z) = zL_F'(z) - L_F(z). \end{aligned} \quad (3.17)$$

The nonincreasing part of condition (3.14) is equivalent to say that

$$L_F(z) \geq 0 \quad \forall z \in (0, +\infty), \quad (3.18)$$

and it is in fact implied by the convexity of F . A simple computation in the case $F \in C^2(0, +\infty)$ shows

$$\frac{d^2}{ds^2} F(s^{-d})s^d = \frac{d^2}{ds^2} \widehat{F}(d \cdot \log s) = \widehat{L}_F^2(d \cdot \log s) \frac{d^2}{s^2} + \widehat{L}_F(d \cdot \log s) \frac{d}{s^2},$$

and therefore

$$(3.14) \text{ is equivalent to } L_F^2(z) \geq -\frac{1}{d}L_F(z) \quad \forall z \in (0, +\infty), \quad (3.19)$$

i.e.,

$$zL_F'(z) \geq \left(1 - \frac{1}{d}\right)L_F(z),$$

the map $z \mapsto z^{1/d-1}L_F(z)$ is nonincreasing. (3.20)

Observe that the bigger is the dimension d , the stronger are the above conditions, which always imply the convexity of F .

REMARK 3.12 (A “dimension free” condition). The weakest condition on F yielding the geodesic convexity of $\mathcal{F}$ in *any dimension* is therefore

$$L_F^2(z) = zL_F'(z) - L_F(z) \geq 0 \quad \forall z \in (0, +\infty). \quad (3.21)$$

Taking into account (3.17), this is also equivalent to ask that

$$\text{the map } s \mapsto F(e^{-s})e^s \text{ is convex and nonincreasing in } (0, +\infty). \quad (3.22)$$

Among the functionals F satisfying (3.14) we quote

$$\text{the entropy functional: } F(s) = s \log s, \quad (3.23)$$

$$\text{the power functional: } F(s) = \frac{1}{m-1}s^m \quad \text{for } m \geq 1 - \frac{1}{d}. \quad (3.24)$$

Observe that the entropy functional and the power functional with $m > 1$ have a superlinear growth. In order to deal with the power functional with $m \leq 1$, due to the failure of the lower semicontinuity property one has to introduce a suitable relaxation $\mathcal{F}^*$ of it, defined by [24,55]

$$\mathcal{F}^*(\mu) := \frac{1}{m-1} \int_{\mathbb{R}^d} u^m(x) d\mathcal{L}^d(x)$$

with $\mu = u \cdot \mathcal{L}^d + \mu_s, \mu_s \perp \mathcal{L}^d$. (3.25)

In this case the functional takes only account of the density of the absolutely continuous part of μ w.r.t. $\mathcal{L}^d$ and the domain of $\mathcal{F}^*$ is the whole $\mathcal{P}_2(\mathbb{R}^d)$. The functional $\mathcal{F}^*$ retains the convexity properties of $\mathcal{F}$, see [9].

EXAMPLE 3.13 (The opposite Wasserstein distance). Let us fix a base measure $\mu^1 \in \mathcal{P}_2(\mathbb{R}^d)$ and let us consider the functional

$$\phi(\mu) := -\frac{1}{2}W_2^2(\mu^1, \mu). \quad (3.26)$$

PROPOSITION 3.14. *For each couple $\mu^2, \mu^3 \in \mathcal{P}_2(\mathbb{R}^d)$ and each transfer plan $\mu^{23} \in \Gamma(\mu^2, \mu^3)$ we have*

$$\begin{aligned} & W_2^2(\mu^1, \mu_t^{2 \rightarrow 3}) \\ & \geq (1-t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) \\ & \quad - t(1-t) \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_2 - x_3|^2 d\mu^{23}(x_2, x_3) \quad \forall t \in [0, 1]. \end{aligned} \quad (3.27)$$

In particular the map $\phi : \mu \mapsto -\frac{1}{2}W_2^2(\mu^1, \mu)$ is (-1) -convex along geodesics.

PROOF. For $\mu^{23} \in \Gamma(\mu^2, \mu^3)$, we can find (see Proposition 7.3.1 of [9]) $\mu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ whose projection on the second and third variable is μ^{23} and such that

$$(\pi^1, (1-t)\pi^2 + t\pi^3)_\# \mu \in \Gamma_0(\mu^1, \mu_t^{2 \rightarrow 3}), \quad (3.28)$$

with $\mu_t^{2 \rightarrow 3} := ((1-t)\pi^2 + t\pi^3)_\# \mu^{23}$. Therefore

$$\begin{aligned} & W_2^2(\mu^1, \mu_t^{2 \rightarrow 3}) \\ & = \int_{\mathbb{R}^{3d}} |(1-t)x_2 + tx_3 - x_1|^2 d\mu(x_1, x_2, x_3) \\ & = \int_{\mathbb{R}^{3d}} ((1-t)|x_2 - x_1|^2 + t|x_3 - x_1|^2 - t(1-t)|x_2 - x_3|^2) d\mu(x_1, x_2, x_3) \\ & \geq (1-t)W_2^2(\mu^1, \mu^2) + tW_2^2(\mu^1, \mu^3) \\ & \quad - t(1-t) \int_{\mathbb{R}^{2d}} |x_2 - x_3|^2 d\mu^{23}(x_2, x_3). \end{aligned} \quad \square$$

In particular, choosing optimal plans in (3.27), we obtain the *semiconcavity inequality* of the Wasserstein distance from a fixed measure μ^3 along the constant speed geodesics $\mu_t^{1 \rightarrow 2}$ connecting μ^1 to μ^2 :

$$\begin{aligned} & W_2^2(\mu_t^{1 \rightarrow 2}, \mu^3) \\ & \geq (1-t)W_2^2(\mu^1, \mu^3) + tW_2^2(\mu^2, \mu^3) - t(1-t)W_2^2(\mu^1, \mu^2). \end{aligned} \quad (3.29)$$

According to Aleksandrov's metric notion of curvature (see [5,58]), this inequality can be interpreted by saying that the Wasserstein space is a positively curved metric space (in short, a *PC-space*). This was already pointed out by a formal computation in [74], showing also that generically the inequality is strict. An example where strict inequality occurs can be obtained as follows: let $d = 2$ and

$$\begin{aligned} \mu^1 &:= \frac{1}{2}(\delta_{(1,1)} + \delta_{(5,3)}), & \mu^2 &:= \frac{1}{2}(\delta_{(-1,1)} + \delta_{(-5,3)}), \\ \mu^3 &:= \frac{1}{2}(\delta_{(0,0)} + \delta_{(0,-4)}). \end{aligned}$$

Then, it is immediate to check that $W_2^2(\mu^1, \mu^2) = 40$, $W_2^2(\mu^1, \mu^3) = 30$ and $W_2^2(\mu^2, \mu^3) = 30$. On the other hand, the unique constant speed geodesic joining μ^1 to μ^2 is given by

$$\mu_t := \frac{1}{2}(\delta_{(1-6t, 1+2t)} + \delta_{(5-6t, 3-2t)})$$

and a simple computation gives

$$24 = W_2^2(\mu_{1/2}, \mu^3) > \frac{30}{2} + \frac{30}{2} - \frac{40}{4}.$$

3.3. Relative entropy and convex functionals of measures

In this section we study in detail the relative entropy functional; although we confine the discussion to a finite-dimensional situation, the formalism used in this section is well adapted to the extension to an infinite-dimensional context, see [9].

DEFINITION 3.15 (Relative entropy). Let γ, μ be Borel probability measures on $\mathbb{R}^d$; the *relative entropy* of μ w.r.t. γ is

$$\mathcal{H}(\mu|\gamma) := \begin{cases} \int_{\mathbb{R}^d} \frac{d\mu}{d\gamma} \log\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise.} \end{cases} \quad (3.30)$$

As in Example 3.8 we introduce the nonnegative, l.s.c. and convex function

$$H(s) := \begin{cases} s(\log s - 1) + 1 & \text{if } s > 0, \\ 1 & \text{if } s = 0, \\ +\infty & \text{if } s < 0, \end{cases} \quad (3.31)$$

and we observe that, whenever $\mu \ll \gamma$, we have

$$\mathcal{H}(\mu|\gamma) = \int_{\mathbb{R}^d} H\left(\frac{d\mu}{d\gamma}\right) d\gamma \geq 0; \quad \mathcal{H}(\mu|\gamma) = 0 \Leftrightarrow \mu = \gamma. \quad (3.32)$$

REMARK 3.16 (Changing γ). Let γ be a Borel measure on $\mathbb{R}^d$ and let $V: \mathbb{R}^d \rightarrow (-\infty, +\infty]$ a Borel map such that

$$\begin{aligned} &V^+ \text{ has at most quadratic growth,} \\ &\tilde{\gamma} := e^{-V} \cdot \gamma \text{ is a probability measure.} \end{aligned} \quad (3.33)$$

Then for measures in $\mathcal{P}_2(\mathbb{R}^d)$ the relative entropy w.r.t. γ is well defined by the formula

$$\mathcal{H}(\mu|\gamma) := \mathcal{H}(\mu|\tilde{\gamma}) - \int_{\mathbb{R}^d} V(x) d\mu(x) \in (-\infty, +\infty] \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d). \quad (3.34)$$

In particular, when γ is the d -dimensional Lebesgue measure, we find the standard entropy functional introduced in (3.23).

More generally, we can consider a

$$\begin{aligned} &\text{proper, l.s.c., convex function } F : [0, +\infty) \rightarrow [0, +\infty] \\ &\text{with superlinear growth} \end{aligned} \quad (3.35)$$

and the related functional

$$\mathcal{F}(\mu|\gamma) := \begin{cases} \int_{\mathbb{R}^d} F\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise.} \end{cases} \quad (3.36)$$

LEMMA 3.17 (Joint lower semicontinuity). *Let $(\gamma^n), (\mu^n) \subset \mathcal{P}(\mathbb{R}^d)$ be two sequences narrowly converging to γ, μ in $\mathcal{P}(\mathbb{R}^d)$. Then*

$$\liminf_{n \rightarrow \infty} \mathcal{H}(\mu^n|\gamma^n) \geq \mathcal{H}(\mu|\gamma), \quad \liminf_{n \rightarrow \infty} \mathcal{F}(\mu^n|\gamma^n) \geq \mathcal{F}(\mu|\gamma). \quad (3.37)$$

The proof of this lemma follows easily from the next representation formula; before stating it, we need to introduce the conjugate function of F

$$F^*(s^*) := \sup_{s \geq 0} s \cdot s^* - F(s) < +\infty \quad \forall s^* \in \mathbb{R}, \quad (3.38)$$

so that

$$F(s) = \sup_{s^* \in \mathbb{R}} s^* \cdot s - F^*(s^*); \quad (3.39)$$

if $s_0 \geq 0$ is a minimizer of F then

$$F^*(s^*) \geq s^* s_0 - F(s_0), \quad s \geq s_0 \implies F(s) = \sup_{s^* \geq 0} s^* \cdot s - F^*(s^*). \quad (3.40)$$

In the case of the entropy functional, we have $H^*(s^*) = e^{s^*} - 1$. Now we recall a classical duality formula for functionals defined on measures; we recall its proof for the reader's convenience.

LEMMA 3.18 (Duality formula). *For any $\gamma, \mu \in \mathcal{P}(\mathbb{R}^d)$ we have*

$$\begin{aligned} &\mathcal{F}(\mu|\gamma) \\ &= \sup \left\{ \int_{\mathbb{R}^d} S^*(x) d\mu(x) - \int_{\mathbb{R}^d} F^*(S^*(x)) d\gamma(x) : S^* \in C_b^0(\mathbb{R}^d) \right\}. \end{aligned} \quad (3.41)$$

PROOF. Up to an addition of a constant, we can always assume $F^*(0) = -\min_{s \geq 0} F(s) = -F(s_0) = 0$. Let us denote by $\mathcal{F}'(\mu|\gamma)$ the right-hand side of (3.41). It is obvious that $\mathcal{F}'(\mu|\gamma) \leq \mathcal{F}(\mu|\gamma)$, so that we have to prove only the converse inequality.

First of all we show that $\mathcal{F}'(\mu|\gamma) < +\infty$ yields that $\mu \ll \gamma$. For, let us fix s^* , $\varepsilon > 0$ and a Borel set A with $\gamma(A) \leq \varepsilon/2$. Since μ, γ are finite measures we can find a compact set $K \subset A$, an open set $G \supset A$ and a continuous function $\zeta : \mathbb{R}^d \rightarrow [0, s^*]$ such that

$$\begin{aligned} \mu(G \setminus K) &\leq \varepsilon, & \gamma(G) &\leq \varepsilon, & \zeta(x) &= s^* \quad \text{on } K, \\ \zeta(x) &= 0 \quad \text{on } \mathbb{R}^d \setminus G. \end{aligned}$$

Since F^* is increasing (by definition (3.38)) and $F^*(0) = 0$, we have

$$\begin{aligned} s^* \mu(K) - F^*(s^*) \varepsilon &\leq \int_K \zeta(x) d\mu(x) - \int_G F^*(\zeta(x)) d\gamma(x) \\ &\leq \int_{\mathbb{R}^d} \zeta(x) d\mu(x) - \int_{\mathbb{R}^d} F^*(\zeta(x)) d\gamma(x) \\ &\leq \mathcal{F}'(\mu|\gamma). \end{aligned}$$

Taking the supremum w.r.t. $K \subset A$ and $s^* \geq 0$, and using (3.40) we get

$$\varepsilon F\left(\frac{\mu(A)}{\varepsilon}\right) \leq \mathcal{F}'(\mu|\gamma) \quad \text{if } \mu(A) \geq \varepsilon s_0.$$

Since $F(s)$ has a superlinear growth as $s \rightarrow +\infty$, we conclude that $\mu(A) \rightarrow 0$ as $\varepsilon \downarrow 0$.

Now we can suppose that $\mu = \rho \cdot \gamma$ for some Borel function $\rho \in L^1(\gamma)$, so that

$$\mathcal{F}'(\mu|\gamma) = \sup \left\{ \int_{\mathbb{R}^d} (S^*(x)\rho(x) - F^*(S^*(x))) d\gamma(x) : S^* \in C_b^0(\mathbb{R}^d) \right\}$$

and, for a suitable dense countable set $C = \{s_n^*\}_{n \in \mathbb{N}} \subset \mathbb{R}$

$$\begin{aligned} \mathcal{F}(\mu|\gamma) &= \int_{\mathbb{R}^d} \sup_{s^* \in C} (s^* \rho(x) - F^*(s^*)) d\gamma(x) \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} \sup_{s^* \in C_k} (s^* \rho(x) - F^*(s^*)) d\gamma(x), \end{aligned}$$

where $C_k = \{s_1^*, \dots, s_k^*\}$. Our thesis follows if we show that for every k ,

$$\int_{\mathbb{R}^d} \max_{s^* \in C_k} (s^* \rho(x) - F^*(s^*)) d\gamma(x) \leq \mathcal{F}'(\mu|\gamma). \quad (3.42)$$

For we call

$$A_j = \{x \in \mathbb{R}^d : s_j^* \rho(x) - F^*(s_j^*) \geq s_i^* \rho(x) - F^*(s_i^*) \quad \forall i \in \{1, \dots, k\}\}$$

and

$$A'_1 = A_1, \quad A'_{j+1} = A_{j+1} \setminus \left(\bigcup_{i=1}^j A_i \right).$$

We find compact sets $K_j \subset A'_j$, open sets $G_j \supset A'_j$ with $G_j \cap K_i = \emptyset$ if $i \neq j$, and continuous functions ζ_j such that

$$\begin{aligned} \sum_{j=1}^k \gamma(G_j \setminus K_j) + \mu(G_j \setminus K_j) &\leq \varepsilon, \\ \zeta_j &\equiv s_j^* \quad \text{on } K_j, \quad \zeta_j \equiv 0 \quad \text{on } \mathbb{R}^d \setminus G_j. \end{aligned}$$

Denoting by $\zeta := \sum_{j=1}^k \zeta_j$, $M := \sum_{j=1}^k |s_j^*|$, since the negative part of $F^*(s^*)$ is bounded above by $|s^*|s_0$ we have

$$\begin{aligned} &\int_{\mathbb{R}^d} \max_{s^* \in C_k} (s^* \rho(x) - F^*(s^*)) \, d\gamma(x) \\ &= \sum_{j=1}^k \int_{A'_j} (s_j^* \rho(x) - F^*(s_j^*)) \, d\gamma(x) \\ &\leq \sum_{j=1}^k \int_{K_j} (s_j^* \rho(x) - F^*(s_j^*)) \, d\gamma(x) + \varepsilon(M + Ms_0) \\ &= \sum_{j=1}^k \int_{K_j} (\zeta(x) \rho(x) - F^*(\zeta(x))) \, d\gamma(x) + \varepsilon(M + Ms_0) \\ &\leq \int_{\mathbb{R}^d} (\zeta(x) \rho(x) - F^*(\zeta(x))) \, d\gamma(x) + \varepsilon(M + Ms_0 + M + F^*(M)). \end{aligned}$$

Passing to the limit as $\varepsilon \downarrow 0$ we get (3.42). □

3.4. Log-concavity and displacement convexity

We want to characterize the probability measures γ inducing a geodesically convex relative entropy functional $\mathcal{H}(\cdot|\gamma)$ in $\mathcal{P}_2(\mathbb{R}^d)$. The following lemma provides the first crucial property; the argument is strictly related to the proof of the Brunn–Minkowski inequality for the Lebesgue measure, obtained via optimal transportation inequalities [86]. See also [18] for the link between log-concavity and representation formulae like (3.50).

LEMMA 3.19 (γ is log-concave if $\mathcal{H}(\cdot|\gamma)$ is displacement convex). *Suppose that for each couple of probability measures $\mu^1, \mu^2 \in \mathcal{P}(\mathbb{R}^d)$ with bounded support there ex-*

ists $\mu \in \Gamma(\mu^1, \mu^2)$ such that $\mathcal{H}(\cdot|\gamma)$ is convex along the interpolating curve $\mu_t^{1 \rightarrow 2} = ((1-t)\pi^1 + t\pi^2)_\# \mu$, $t \in [0, 1]$. Then for each couple of open sets $A, B \subset \mathbb{R}^d$ and $t \in [0, 1]$ we have

$$\log \gamma((1-t)A + tB) \geq (1-t) \log \gamma(A) + t \log \gamma(B). \quad (3.43)$$

PROOF. We can obviously assume that $\gamma(A) > 0$, $\gamma(B) > 0$ in (3.43); we consider

$$\mu^1 := \gamma(\cdot|A) = \frac{1}{\gamma(A)} \chi_A \cdot \gamma, \quad \mu^2 := \gamma(\cdot|B) = \frac{1}{\gamma(B)} \chi_B \cdot \gamma,$$

observing that

$$\mathcal{H}(\mu^1|\gamma) = -\log \gamma(A), \quad \mathcal{H}(\mu^2|\gamma) = -\log \gamma(B). \quad (3.44)$$

If $\mu_t^{1 \rightarrow 2}$ is induced by a transfer plan $\mu \in \Gamma(\mu^1, \mu^2)$ along which the relative entropy is displacement convex, we have

$$\mathcal{H}(\mu_t^{1 \rightarrow 2}|\gamma) \leq (1-t)\mathcal{H}(\mu^1|\gamma) + t\mathcal{H}(\mu^2|\gamma) = -(1-t)\log \gamma(A) - t\log \gamma(B).$$

On the other hand, the measure $\mu_t^{1 \rightarrow 2}$ is concentrated on $(1-t)A + tB = \pi_t^{1 \rightarrow 2}(A \times B)$ and the next lemma shows that

$$-\log \gamma((1-t)A + tB) \leq \mathcal{H}(\mu_t^{1 \rightarrow 2}|\gamma). \quad \square$$

LEMMA 3.20 (Relative entropy of concentrated measures). *Let $\gamma, \mu \in \mathcal{P}(\mathbb{R}^d)$; if μ is concentrated on a Borel set A , i.e., $\mu(\mathbb{R}^d \setminus A) = 0$, then*

$$\mathcal{H}(\mu|\gamma) \geq -\log \gamma(A). \quad (3.45)$$

PROOF. It is not restrictive to assume $\mu \ll \gamma$ and $\gamma(A) > 0$; denoting by γ_A the probability measure $\gamma(\cdot|A) := \gamma(A)^{-1} \chi_A \cdot \gamma$, we have

$$\begin{aligned} \mathcal{H}(\mu|\gamma) &= \int_{\mathbb{R}^d} \log\left(\frac{d\mu}{d\gamma}\right) d\mu \\ &= \int_A \log\left(\frac{d\mu}{d\gamma_A} \cdot \frac{1}{\gamma(A)}\right) d\mu \\ &= \int_A \log\left(\frac{d\mu}{d\gamma_A}\right) d\mu - \int_A \log(\gamma(A)) d\mu \\ &= \mathcal{H}(\mu|\gamma_A) - \log(\gamma(A)) \\ &\geq -\log(\gamma(A)). \end{aligned} \quad \square$$

The previous results justifies the following definition.

DEFINITION 3.21 (log-concavity of a measure). We say that a Borel probability measure $\gamma \in \mathcal{P}(\mathbb{R}^d)$ is *log-concave* if for every couple of open sets $A, B \subset \mathbb{R}^d$, we have

$$\log \gamma((1-t)A + tB) \geq (1-t) \log \gamma(A) + t \log \gamma(B). \quad (3.46)$$

In Definition 3.21 and also in the previous theorem we confined ourselves to pairs of open sets, to avoid the nontrivial issue of the measurability of $(1-t)A + tB$ when A and B are only Borel (in fact, it is an open set whenever A and B are open). Observe that a log-concave measure γ in particular satisfies

$$\log \gamma(B_r((1-t)x_0 + tx_1)) \geq (1-t) \log \gamma(B_r(x_0)) + t \log \gamma(B_r(x_1)), \quad (3.47)$$

for every couple of points $x_0, x_1 \in \mathbb{R}^d$, $r > 0$, $t \in [0, 1]$.

We want to show that in fact log-concavity is equivalent to the geodesic convexity of the relative entropy functional $\mathcal{H}(\cdot|\gamma)$.

Let us first recall some elementary properties of convex sets in $\mathbb{R}^d$. Let $C \subset \mathbb{R}^d$ be a convex set; the *affine dimension* $\dim C$ of C is the linear dimension of its affine envelope

$$\text{aff } C = \{(1-t)x_0 + tx_1 : x_0, x_1 \in C, t \in \mathbb{R}\}, \quad (3.48)$$

which is an affine subspace of $\mathbb{R}^d$. We denote by $\text{int } C$ the relative interior of C as a subset of $\text{aff } C$: it is possible to show that

$$\text{int } C \neq \emptyset, \quad \overline{\text{int } C} = \overline{C}, \quad \mathcal{H}^k(\overline{C} \setminus \text{int } C) = 0 \quad \text{if } k = \dim C, \quad (3.49)$$

where $\mathcal{H}^k$ is the k -dimensional Hausdorff measure in $\mathbb{R}^d$. The previous theorem shows that log-concavity of γ is equivalent to the convexity of $\mathcal{H}(\mu|\gamma)$ along geodesics of the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$: the link between these two concepts is provided by the representation formula (3.50).

THEOREM 3.22. *Let us suppose that $\gamma \in \mathcal{P}(\mathbb{R}^d)$ satisfies the log-concavity assumptions on balls (3.47). Then $\text{supp } \gamma$ is convex and there exists a convex l.s.c. function $V : \mathbb{R}^d \rightarrow (\infty, +\infty]$ such that*

$$\gamma = e^{-V} \mathcal{H}^k|_{\text{aff}(\text{supp } \gamma)}, \quad \text{where } k = \dim(\text{supp } \gamma). \quad (3.50)$$

Conversely, if γ admits the representation (3.50) then γ is log-concave and the relative entropy functional $\mathcal{H}(\cdot|\gamma)$ is convex along any geodesic of $\mathcal{P}_2(\mathbb{R}^d)$.

PROOF. Let us suppose that γ satisfies the log-concavity inequality on balls and let k be the dimension of $\text{aff}(\text{supp } \gamma)$. Observe that the measure γ satisfies the same inequality (3.47) for the balls of $\text{aff}(\text{supp } \gamma)$: up to an isometric change of coordinates it is not restrictive to assume that $k = d$ and $\text{aff}(\text{supp } \gamma) = \mathbb{R}^d$.

Let us now introduce the set

$$D := \left\{ x \in \mathbb{R}^d : \liminf_{r \downarrow 0} \frac{\gamma(B_r(x))}{r^d} > 0 \right\}. \quad (3.51)$$

Since (3.47) yields

$$\frac{\gamma(B_r(x_t))}{r^d} \geq \left(\frac{\gamma(B_r(x_0))}{r^d} \right)^{1-t} \left(\frac{\gamma(B_r(x_1))}{r^d} \right)^t, \quad t \in (0, 1), \quad (3.52)$$

it is immediate to check that D is a convex subset of $\mathbb{R}^d$ with $D \subset \text{supp } \gamma$.

General results on derivation of Radon measures in $\mathbb{R}^d$ (see, for instance, Theorem 2.56 in [8]) show that

$$\limsup_{r \downarrow 0} \frac{\gamma(B_r(x))}{r^d} < +\infty \quad \text{for } \mathcal{L}^d\text{-a.e. } x \in \mathbb{R}^d \quad (3.53)$$

and

$$\limsup_{r \downarrow 0} \frac{r^d}{\gamma(B_r(x))} < +\infty \quad \text{for } \gamma\text{-a.e. } x \in \mathbb{R}^d. \quad (3.54)$$

Using (3.54) we see that actually γ is concentrated on D (so that $\text{supp } \gamma \subset \overline{D}$) and therefore, being d the dimension of $\text{aff}(\text{supp } \gamma)$, it follows that d is also the dimension of $\text{aff}(D)$.

If a point $\bar{x} \in \mathbb{R}^d$ exists such that

$$\limsup_{r \downarrow 0} \frac{\gamma(B_r(\bar{x}))}{r^d} = +\infty,$$

then (3.52) forces every point of $\text{int}(D)$ to verify the same property, but this would be in contradiction with (3.53), since we know that $\text{int}(D)$ has strictly positive $\mathcal{L}^d$ -measure. Therefore

$$\limsup_{r \downarrow 0} \frac{\gamma(B_r(x))}{r^d} < +\infty \quad \text{for all } x \in \mathbb{R}^d, \quad (3.55)$$

and we obtain that $\gamma \ll \mathcal{L}^d$, again by the theory of derivation of Radon measures in $\mathbb{R}^d$. In the sequel we denote by g the density of γ w.r.t. $\mathcal{L}^d$ and notice that by Lebesgue differentiation theorem $g > 0$ $\mathcal{L}^d$ -a.e. in D and $g = 0$ $\mathcal{L}^d$ -a.e. in $\mathbb{R}^d \setminus D$.

By (3.47) the maps

$$V_r(x) = -\log\left(\frac{\gamma(B_r(x))}{\omega_d r^d}\right)$$

are convex on $\mathbb{R}^d$, and (3.55) gives that the family $V_r(x)$ is bounded as $r \downarrow 0$ for any $x \in D$. Using the pointwise boundedness of V_r on D and the convexity of V_r it is easy to show that

V_r are locally equibounded (hence locally equicontinuous) on $\text{int}(D)$ as $r \downarrow 0$. Let W be a limit point of V_r , with respect to the local uniform convergence, as $r \downarrow 0$: W is convex on $\text{int}(D)$ and Lebesgue differentiation theorem shows that

$$\lim_{r \downarrow 0} V_r(x) = -\log g(x) = W(x) \quad \text{for } \mathcal{L}^d\text{-a.e. } x \in \text{int}(D), \quad (3.56)$$

so that $\gamma = g \mathcal{L}^d = e^{-W} \chi_{\text{int}(D)} \mathcal{L}^d$. In order to get a globally defined convex and l.s.c. function V we extend W with the $+\infty$ value out of $\text{int}(D)$ and define V to be its convex and l.s.c. envelope. It turns out that V coincides with W on $\text{int}(D)$, so that still the representation $\gamma = e^{-V} \mathcal{L}^d$ holds.

Conversely, let us suppose that γ admits the representation (3.50) for a given convex l.s.c. function V and let $\mu^1, \mu^2 \in \mathcal{P}_2(\mathbb{R}^d)$; if their relative entropies are finite then they are absolutely continuous w.r.t. γ and therefore their supports are contained in $\text{aff}(\text{supp } \gamma)$. It follows that the support of any optimal plan $\mu \in \Gamma_o(\mu^1, \mu^2)$ in $\mathcal{P}_2(\mathbb{R}^d)$ is contained in $\text{aff}(\text{supp } \gamma) \times \text{aff}(\text{supp } \gamma)$: up to a linear isometric change of coordinates, it is not restrictive to suppose $\text{aff}(\text{supp } \gamma) = \mathbb{R}^d$, $\mu^1, \mu^2 \in \mathcal{P}_2^a(\mathbb{R}^d)$, $\gamma = e^{-V} \mathcal{L}^d \in \mathcal{P}(\mathbb{R}^d)$.

In this case we introduce the densities u^i of μ^i w.r.t. $\mathcal{L}^d$, observing that

$$\frac{d\mu^i}{d\gamma} = u^i e^V, \quad i = 1, 2,$$

where we adopted the convention $0 \cdot (+\infty) = 0$ (recall that $u^i(x) = 0$ for $\mathcal{L}^d$ -a.e. $x \in \mathbb{R}^d \setminus D(V)$). Therefore the entropy functional can be written as

$$\mathcal{H}(\mu^i | \gamma) = \int_{\mathbb{R}^d} u^i(x) \log u^i(x) dx + \int_{\mathbb{R}^d} V(x) d\mu^i(x), \quad (3.57)$$

i.e., the sum of two geodesically convex functionals, as we proved discussing Examples 3.4 and 3.8. Lemma 3.19 yields the log-concavity of γ . $\square$

If γ is log-concave and F satisfies (3.22), then all the integral functionals $\mathcal{F}(\cdot | \gamma)$ introduced in (3.36) are geodesically convex in $\mathcal{P}_2(\mathbb{R}^d)$.

THEOREM 3.23 (Geodesic convexity for relative integral functionals). *Suppose that γ is log-concave and $F : [0, +\infty) \rightarrow [0, +\infty]$ satisfies conditions (3.35) and (3.22). Then the integral functional $\mathcal{F}(\cdot | \gamma)$ is geodesically convex in $\mathcal{P}_2(\mathbb{R}^d)$.*

PROOF. Arguing as in the final part of the proof of Theorem 3.22 we can assume that $\gamma := e^{-V} \mathcal{L}^d$ for a convex l.s.c. function $V : \mathbb{R}^d \rightarrow (-\infty, +\infty]$ whose domain has not empty interior. For every couple of measures $\mu^1, \mu^2 \in D(\mathcal{F}(\cdot | \gamma))$ we have

$$\mu^i = u^i e^V \cdot \gamma, \quad \mathcal{F}(\mu^i | \gamma) = \int_{\mathbb{R}^d} F(u^i(x) e^{V(x)}) e^{-V(x)} dx, \quad i = 1, 2. \quad (3.58)$$

We denote by $\mathbf{r}$ the optimal transport map for the Wasserstein distance pushing μ^1 to μ^2 and we set $\mathbf{r}_t := (1-t)\mathbf{i} + t\mathbf{r}$, $\mu_t := (\mathbf{r}_t)_\# \mu^1$; arguing as in Proposition 3.11, we get

$$\mathcal{F}(\mu_t|\gamma) = \int_{\mathbb{R}^d} F\left(\frac{u(x)e^{V(\mathbf{r}_t(x))}}{\det \nabla \mathbf{r}_t(x)}\right) \det \nabla \mathbf{r}_t(x) e^{-V(\mathbf{r}_t(x))} dx, \quad (3.59)$$

and the integrand above may be seen as the composition of the convex and nonincreasing map $s \mapsto F(u(x)e^{-s})e^s$ with the concave curve

$$t \mapsto -V(\mathbf{r}_t(x)) + \log(\det \nabla \mathbf{r}_t(x)),$$

since $D(x) := \nabla \mathbf{r}(x)$ is a diagonalizable map with nonnegative eigenvalues and

$$t \mapsto \log \det((1-t)I + tD(x)) \quad \text{is concave in } [0, 1]. \quad \square$$

4. Subdifferential calculus in $\mathcal{P}_2(\mathbb{R}^d)$

Let X be an Hilbert space. In the classical theory of subdifferential calculus (see, e.g., [22]) for lower semicontinuous functionals $\phi : X \rightarrow (-\infty, +\infty]$ with proper domain

$$D(\phi) := \{v \in X : \phi(v) < +\infty\} \neq \emptyset, \quad (4.1)$$

the *Fréchet subdifferential* $\partial\phi : X \rightarrow 2^X$ of ϕ is a multivalued operator defined as

$$\begin{aligned} \xi \in \partial\phi(v) &\iff v \in D(\phi), \\ \liminf_{w \rightarrow v} \frac{\phi(w) - \phi(v) - \langle \xi, w - v \rangle}{|w - v|} &\geq 0, \end{aligned} \quad (4.2)$$

which we will also write in the equivalent form for $v \in D(\phi)$

$$\begin{aligned} \xi \in \partial\phi(v) &\iff \phi(w) \geq \phi(v) + \langle \xi, w - v \rangle + o(|w - v|) \\ &\text{as } w \rightarrow v. \end{aligned} \quad (4.3)$$

As usual in multivalued analysis, the proper domain $D(\partial\phi) \subset D(\phi)$ is defined as the set of all $v \in X$ such that $\partial\phi(v) \neq \emptyset$; we will use this convention for all the multivalued operators we will introduce.

The metric counterpart of the Fréchet subdifferential is represented by the *metric slope* of ϕ , which for every $v \in D(\phi)$ is defined by

$$|\partial\phi|(v) = \limsup_{w \rightarrow v} \frac{(\phi(v) - \phi(w))^+}{|w - v|}, \quad (4.4)$$

and can also be characterized by an asymptotic expansion similar to (4.3) for $s \geq 0$

$$\begin{aligned} s \geq |\partial\phi|(v) &\iff \phi(w) \geq \phi(v) - s|w - v| + o(|w - v|) \\ &\text{as } w \rightarrow v. \end{aligned} \quad (4.5)$$

It is then immediate to check that

$$\xi \in \partial\phi(v) \implies |\partial\phi|(v) \leq |\xi|. \quad (4.6)$$

The Fréchet subdifferential and the metric slope occur quite naturally in the Euler equations for minima of (smooth perturbation of) ϕ .

A. Euler equation for quadratic perturbations. If v_τ is a minimizer of

$$w \mapsto \Phi(\tau, v; w) := \phi(w) + \frac{1}{2\tau}|w - v|^2 \quad \text{for some } \tau > 0, v \in X \quad (4.7)$$

then

$$v_\tau \in D(\partial\phi) \quad \text{and} \quad -\frac{v_\tau - v}{\tau} \in \partial\phi(v_\tau); \quad (4.8)$$

concerning the slope we easily get

$$v_\tau \in D(|\partial\phi|) \quad \text{and} \quad |\partial\phi|(v) \leq \frac{|v - v_\tau|}{\tau}. \quad (4.9)$$

For λ -convex functionals the Fréchet subdifferential enjoys at least two other simple but fundamental properties, which play a crucial role in the corresponding variational theory of evolution equations.

B. Characterization by variational inequalities and monotonicity. If ϕ is λ -convex, then

$$\begin{aligned} \xi \in \partial\phi(v) &\iff \phi(w) \geq \phi(v) + \langle \xi, w - v \rangle + \frac{\lambda}{2}|w - v|^2 \\ &\forall w \in D(\phi); \end{aligned} \quad (4.10)$$

in particular,

$$\begin{aligned} \xi_i \in \partial\phi(v_i) &\implies \langle \xi_1 - \xi_2, v_1 - v_2 \rangle \geq \lambda|v_1 - v_2|^2 \\ &\forall v_1, v_2 \in D(\partial\phi). \end{aligned} \quad (4.11)$$

As in (4.10), the slope of a λ -convex functional can also be characterized by a system of inequalities for $s \geq 0$

$$\begin{aligned} s \geq |\partial\phi|(v) &\iff \phi(w) \geq \phi(v) - s|w - v| + \frac{\lambda}{2}|w - v|^2 \\ &\forall w \in D(\phi), \end{aligned} \quad (4.12)$$

which can equivalently reformulated as

$$|\partial\phi|(v) = \sup_{w \neq v} \left(\frac{\phi(v) - \phi(w)}{|v - w|} + \frac{\lambda}{2} |v - w| \right)^+. \quad (4.13)$$

C. Convexity and strong-weak closure ([22], Chapter II, Example 2.3.4, Proposition 2.5). If ϕ is λ -convex, then $\partial\phi(v)$ is closed and convex, and for every sequences $(v_n) \subset X$, $(\xi_n) \subset X$ we have

$$\xi_n \in \partial\phi(v_n), \quad v_n \rightarrow v, \xi_n \rightharpoonup \xi \implies \xi \in \partial\phi(v), \quad \phi(v_n) \rightarrow \phi(v). \quad (4.14)$$

The slope is l.s.c.

$$v_n \rightarrow v \implies \liminf_{n \rightarrow \infty} |\partial\phi|(v_n) \geq |\partial\phi|(v). \quad (4.15)$$

Modeled on the last property **C**, and following a terminology introduced by Clarke, see, e.g., [80], Chapter 8, we say that a functional ϕ is *regular* if

$$\begin{cases} \xi_n \in \partial\phi(v_n), & \varphi_n = \phi(v_n) \\ v_n \rightarrow v, & \xi_n \rightharpoonup \xi, \quad \varphi_n \rightarrow \varphi \end{cases} \implies \xi \in \partial\phi(v), \quad \varphi = \phi(v). \quad (4.16)$$

D. Minimal selection and slope. If ϕ is regular (in particular if ϕ is λ -convex) $|\partial\phi|(v)$ is finite if and only if $\partial\phi(v) \neq \emptyset$ and

$$|\partial\phi|(v) = \min\{|\xi| : \xi \in \partial\phi(v)\}. \quad (4.17)$$

The inequality $\leq$ in (4.17) follows directly from (4.6). The other one is simple to check, using the Hahn–Banach theorem, in the λ -convex case. In the more general case when ϕ is regular, one can use the existence (proved even in a general metric setting in Lemma 3.1.5 of [9]) of an infinitesimal sequence $(\tau_n) \subset (0, +\infty)$ and minimizers v_n of $w \mapsto \phi(w) + |w - v|^2/2\tau_n$ such that $\phi(v_n) \rightarrow \phi(v)$ and

$$\lim_{n \rightarrow \infty} \frac{|v - v_n|}{\tau_n} = |\partial\phi|(v).$$

As $(v - v_n)/\tau_n \in \partial\phi(v_n)$ we can use the regularity property and a weak compactness argument to obtain $\xi \in \partial\phi(v)$ with $|\xi| \leq |\partial\phi|(v)$.

E. Chain rule. If $v : (a, b) \rightarrow D(\phi)$ is a curve in X then

$$\frac{d}{dt} \phi(v(t)) = \langle \xi, v'(t) \rangle \quad \forall \xi \in \partial\phi(v(t)), \quad (4.18)$$

at each point t where v and $\phi \circ v$ are differentiable and $\partial\phi(v(t)) \neq \emptyset$. In particular (see [22], Chapter III, Lemma 3.3, and Corollary 2.4.10 in [9]) if ϕ is also λ -convex, $v \in AC(a, b; X)$,

and

$$\int_a^b |\partial\phi|(v(t))|v'(t)| dt < +\infty, \quad (4.19)$$

then $\phi \circ v$ is absolutely continuous in (a, b) and (4.18) holds for $\mathcal{L}^1$ -a.e. $t \in (a, b)$.

The aim of this section is to extend the notion of Fréchet subdifferentiability and these properties to the Wasserstein framework (see also [29] for related results).

4.1. Definition of the subdifferential for a.c. measures

In this section we focus our attention to functionals ϕ defined on $\mathcal{P}_2(\mathbb{R}^d)$. The formal mechanism for translating statements from the euclidean framework to the Wasserstein formalism is simple: if $\mu \leftrightarrow v$ is the reference point, scalar products $\langle \cdot, \cdot \rangle$ have to be intended in the reference Hilbert space $L^2(\mu; \mathbb{R}^d)$ (which contains the tangent space $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$) and displacement vectors $w - v$ corresponds to transport maps $\mathbf{t}_\mu^v - \mathbf{i}$, which are well defined if $\mu \in \mathcal{P}_2^a(\mathbb{R}^d)$. According to these two natural rules, the transposition of (4.2) yields:

DEFINITION 4.1 (Fréchet subdifferential and metric slope). Let us consider a functional $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ and a measure $\mu \in D(\phi) \cap \mathcal{P}_2^a(\mathbb{R}^d)$. We say that $\xi \in L^2(\mu; \mathbb{R}^d)$ belongs to the Fréchet subdifferential $\partial\phi(\mu)$ if

$$\phi(v) - \phi(\mu) \geq \int_{\mathbb{R}^d} \langle \xi(x), \mathbf{t}_\mu^v(x) - x \rangle d\mu(x) + o(W_2(\mu, v)). \quad (4.20)$$

When $\xi \in \partial\phi(\mu)$ also satisfies

$$\phi(\mathbf{t}_\# \mu) - \phi(\mu) \geq \int_{\mathbb{R}^d} \langle \xi(x), \mathbf{t}(x) - x \rangle d\mu(x) + o(\|\mathbf{t} - \mathbf{i}\|_{L^2(\mu; \mathbb{R}^d)}), \quad (4.21)$$

then we will say that ξ is a *strong* subdifferential.

It is obvious that $\partial\phi(\mu)$ is a closed convex subset of $L^2(\mu; \mathbb{R}^d)$; in fact, we could also impose that it is contained in the tangent space $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$, since the vector ξ in (4.20) acts only on tangent vectors (see Theorem 2.22): for, if Π denotes the orthogonal projection onto $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ in $L^2(\mu; \mathbb{R}^d)$,

$$\xi \in \partial\phi(\mu) \implies \Pi\xi \in \partial\phi(\mu). \quad (4.22)$$

It is interesting to note that elements in $\partial\phi(\mu) \cap \text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ are in fact *strong* subdifferentials.

PROPOSITION 4.2 (Subdifferentials in $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ are strong). *Let $\mu \in D(\phi) \cap \mathcal{P}_2^a(\mathbb{R}^d)$ and let $\xi \in \partial\phi(\mu) \cap \text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$. Then ξ is a strong subdifferential.*

PROOF. We argue by contradiction, and we assume that a constant $\delta > 0$ and a sequence $(\mathbf{s}_n) \subset L^2(\mu; \mathbb{R}^d)$ with $\varepsilon_n := \|\mathbf{s}_n - \mathbf{i}\|_{L^2(\mu; \mathbb{R}^d)} \rightarrow 0$ as $n \rightarrow \infty$ exist such that

$$\phi(\mu_n) - \phi(\mu) - \int_{\mathbb{R}^d} \langle \xi, \mathbf{s}_n - \mathbf{i} \rangle d\mu \leq -\delta \varepsilon_n, \quad \mu_n := (\mathbf{s}_n)_\# \mu. \quad (4.23)$$

Let us denote by $\mathbf{t}_n$ the optimal transport pushing μ onto μ_n : we know that

$$\|\mathbf{t}_n - \mathbf{i}\|_{L^2(\mu; \mathbb{R}^d)} = W_2(\mu, \mu_n) \leq \varepsilon_n \rightarrow 0. \quad (4.24)$$

By the definition of subdifferential, there exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$

$$\phi(\mu_n) - \phi(\mu) \geq \int_{\mathbb{R}^d} \langle \xi, \mathbf{t}_n - \mathbf{i} \rangle d\mu - \frac{\delta}{2} \varepsilon_n;$$

combining with (4.23) we obtain

$$\int_{\mathbb{R}^d} \langle \xi, \mathbf{t}_n - \mathbf{s}_n \rangle d\mu \leq -\frac{\delta}{2} \varepsilon_n \quad \forall n \geq n_0. \quad (4.25)$$

Up to an extraction of a suitable subsequence, we can assume that

$$\frac{\mathbf{s}_n - \mathbf{i}}{\varepsilon_n} \rightharpoonup \tilde{\mathbf{s}}, \quad \frac{\mathbf{t}_n - \mathbf{i}}{\varepsilon_n} \rightharpoonup \tilde{\mathbf{t}} \quad \text{weakly in } L^2(\mu; \mathbb{R}^d) \text{ as } n \rightarrow \infty; \quad (4.26)$$

by (4.25) we get

$$\int_{\mathbb{R}^d} \langle \xi, \tilde{\mathbf{t}} - \tilde{\mathbf{s}} \rangle d\mu \leq -\frac{\delta}{2} < 0. \quad (4.27)$$

On the other hand, for every function $\zeta \in C_c^\infty(\mathbb{R}^d)$, the global estimates

$$\begin{aligned} \zeta(y) - \zeta(x) &\leq \langle D\zeta(x), y - x \rangle + C|y - x|^2, \\ \zeta(x) - \zeta(y) &\leq \langle D\zeta(x), x - y \rangle + C|y - x|^2 \end{aligned}$$

for some constant $C \geq 0$ yield

$$\begin{aligned} 0 &= \int_{\mathbb{R}^d} (\zeta(\mathbf{t}_n(x)) - \zeta(\mathbf{s}_n(x))) d\mu(x) \\ &\leq \int_{\mathbb{R}^d} \langle D\zeta(x), \mathbf{t}_n(x) - \mathbf{s}_n(x) \rangle d\mu(x) \\ &\quad + C \int_{\mathbb{R}^d} (|\mathbf{s}_n(x) - x|^2 + |\mathbf{t}_n(x) - x|^2) d\mu(x) \\ &\leq \int_{\mathbb{R}^d} \langle D\zeta(x), \mathbf{t}_n(x) - \mathbf{s}_n(x) \rangle d\mu(x) + 2C\varepsilon_n^2. \end{aligned}$$

Dividing by ε_n and passing to the limit as $n \rightarrow \infty$ we get

$$\int_{\mathbb{R}^d} \langle D\zeta, \tilde{\mathbf{t}} - \tilde{\mathbf{s}} \rangle d\mu \geq 0 \quad \forall \zeta \in C_c^\infty(\mathbb{R}^d). \quad (4.28)$$

Since the gradients of $C_c^\infty(\mathbb{R}^d)$ functions are dense in $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$, (4.28) contradicts (4.27). $\square$

The De Giorgi's definition of the metric slope of ϕ is in fact common to functionals defined in arbitrary metric spaces [37].

DEFINITION 4.3 (Metric slope). Let us consider a functional $\phi: \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ and a measure $\mu \in D(\phi)$. The metric slope of ϕ at μ is defined by

$$|\partial\phi|(\mu) = \limsup_{\nu \rightarrow \mu} \frac{(\phi(\mu) - \phi(\nu))^+}{W_2(\nu, \mu)}, \quad (4.29)$$

or, equivalently, by

$$|\partial\phi|(\mu) := \inf \left\{ s \geq 0: \phi(\nu) \geq \phi(\mu) - sW_2(\nu, \mu) + o(W_2(\nu, \mu)) \right. \\ \left. \text{as } W_2(\nu, \mu) \rightarrow 0 \right\}. \quad (4.30)$$

4.2. Subdifferential calculus in $\mathcal{P}_2^a(\mathbb{R}^d)$

We now try to reproduce in the Wasserstein framework the calculus properties for the subdifferential, we briefly discussed at the beginning of the present section.

In order to simplify some technical point, we are supposing that

$$\phi: \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty] \text{ is proper and lower semicontinuous} \\ \text{with } D(|\partial\phi|) \subset \mathcal{P}_2^a(\mathbb{R}^d), \quad (4.31a)$$

and that for some $\tau_* > 0$ the functional

$$\nu \mapsto \Phi(\tau, \mu; \nu) = 1/(2\tau)W_2^2(\mu, \nu) + \phi(\nu) \text{ admits at least} \\ \text{a minimum point } \mu_\tau \text{ for all } \tau \in (0, \tau_*) \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d). \quad (4.31b)$$

Notice that $D(\phi) \subset \mathcal{P}_2^a(\mathbb{R}^d)$ is a sufficient but not necessary condition for (4.31a): the internal energy functionals induced by a class of sublinear functions F satisfy (4.31a), but have a domain strictly larger than $\mathcal{P}_2^a(\mathbb{R}^d)$ (see Theorem 10.4.8 of [9]).

A. Euler equation for quadratic perturbations. When we want to minimize the perturbed functional (4.31b) we get a result completely analogous to the Euclidean one:

LEMMA 4.4. *Let ϕ be satisfying (4.31a,b). Each minimizer μ_τ of (4.31b) belongs to $D(|\partial\phi|)$ and*

$$\frac{1}{\tau}(\mathbf{t}_{\mu_\tau}^\mu - \mathbf{i}) \in \partial\phi(\mu_\tau) \quad \text{is a strong subdifferential.} \quad (4.32)$$

PROOF. The minimality of μ_τ gives for every $\nu \in \mathcal{P}_2(\mathbb{R}^d)$

$$\begin{aligned} \phi(\nu) - \phi(\mu_\tau) &= \Phi(\tau, \mu; \nu) - \Phi(\tau, \mu; \mu_\tau) \\ &\quad + \frac{1}{2\tau}(W_2^2(\mu_\tau, \mu) - W_2^2(\nu, \mu)) \\ &\geq \frac{1}{2\tau}(W_2^2(\mu_\tau, \mu) - W_2^2(\nu, \mu)) \end{aligned} \quad (4.33)$$

$$\geq -\frac{1}{2\tau}W_2(\mu_\tau, \nu)(W_2(\mu_\tau, \mu) + W_2(\nu, \mu)). \quad (4.34)$$

Letting ν converge to μ_τ , (4.34) yields

$$|\partial\phi|(\mu_\tau) \leq \frac{W_2(\mu_\tau, \nu)}{\tau}. \quad (4.35)$$

By (4.31a) we get $\mu_\tau \in \mathcal{P}_2^a(\mathbb{R}^d)$; if $\nu = \mathbf{t}_\# \mu_\tau$ we have

$$\begin{aligned} W_2^2(\mu_\tau, \mu) &= \int_{\mathbb{R}^d} |\mathbf{t}_{\mu_\tau}^\mu(x) - x|^2 d\mu_\tau(x), \\ W_2^2(\nu, \mu) &\leq \int_{\mathbb{R}^d} |\mathbf{t}(x) - \mathbf{t}_{\mu_\tau}^\mu(x)|^2 d\mu_\tau(x), \end{aligned}$$

and therefore the elementary identity $\frac{1}{2}|a|^2 - \frac{1}{2}|b|^2 = \langle a, a - b \rangle - \frac{1}{2}|a - b|^2$ and (4.33) yield

$$\begin{aligned} \phi(\nu) - \phi(\mu_\tau) &\geq \frac{1}{2\tau} \int_{\mathbb{R}^d} (|\mathbf{t}_{\mu_\tau}^\mu(x) - x|^2 - |\mathbf{t}_{\mu_\tau}^\mu(x) - \mathbf{t}(x)|^2) d\mu_\tau(x) \\ &= \int_{\mathbb{R}^d} \left(\frac{1}{\tau} \langle \mathbf{t}_{\mu_\tau}^\mu(x) - x, \mathbf{t}(x) - x \rangle - \frac{1}{2\tau} |\mathbf{t}(x) - x|^2 \right) d\mu_\tau(x) \\ &= \int_{\mathbb{R}^d} \frac{1}{\tau} \langle \mathbf{t}_{\mu_\tau}^\mu(x) - x, \mathbf{t}(x) - x \rangle d\mu_\tau(x) - \frac{1}{2\tau} \|\mathbf{t} - \mathbf{i}\|_{L^2(\mu_\tau; \mathbb{R}^d)}^2. \end{aligned}$$

We deduce $1/\tau(\mathbf{t}_{\mu_\tau}^\mu - \mathbf{i}) \in \partial\phi(\mu_\tau)$ and the strong subdifferentiability condition. $\square$

The above result, though simple, is very useful and usually provides the first crucial information when one looks for the properties of solutions of the variational problem (4.31b).

The nice argument which combines the minimality of μ_τ and the possibility to use any “test” transport map $\mathbf{t}$ to estimate $W_2^2(\mathbf{t}_\# \nu, \mu)$ was originally introduced by Otto.

4.3. The case of λ -convex functionals along geodesics

Let us now focus our attention to the case of a λ -convex functional:

$$\phi \text{ is } \lambda\text{-convex on geodesics, according to Definition 3.1.} \quad (4.36)$$

B. Characterization by variational inequalities and monotonicity. Suppose that ϕ satisfies (4.31a,b) and (4.36). Then a vector $\xi \in L^2(\mu; \mathbb{R}^d)$ belongs to the Fréchet subdifferential of ϕ at μ iff

$$\begin{aligned} & \phi(\nu) - \phi(\mu) \\ & \geq \int_{\mathbb{R}^d} \langle \xi(x), \mathbf{t}_\mu^\nu(x) - x \rangle d\mu(x) + \frac{\lambda}{2} W_2^2(\mu, \nu) \quad \forall \nu \in D(\phi). \end{aligned} \quad (4.37)$$

In particular if $\xi_i \in \partial\phi(\mu_i)$, $i = 1, 2$, and $\mathbf{t} = \mathbf{t}_{\mu_1}^{\mu_2}$ is the optimal transport map, then

$$\int_{\mathbb{R}^d} \langle \xi_2(\mathbf{t}(x)) - \xi_1(x), \mathbf{t}(x) - x \rangle d\mu_1(x) \geq \lambda W_2^2(\mu_1, \mu_2). \quad (4.38)$$

Concerning the slope of ϕ we have for every $s \geq 0$

$$\begin{aligned} s \geq |\partial\phi|(\mu) & \iff \phi(\nu) \geq \phi(\mu) - s W_2(\nu, \mu) + \frac{\lambda}{2} W_2^2(\nu, \mu) \\ & \forall \nu \in D(\phi), \end{aligned} \quad (4.39)$$

or, equivalently,

$$|\partial\phi|(\mu) = \sup_{\nu \neq \mu} \left(\frac{\phi(\mu) - \phi(\nu)}{W_2(\mu, \nu)} + \frac{\lambda}{2} W_2(\mu, \nu) \right)^+. \quad (4.40)$$

PROOF. One implication of (4.37) and of (4.39) is trivial. To prove the other one, in the case of (4.37) suppose that $\xi \in \partial\phi(\mu)$ and $\nu \in D(\phi)$; for $t \in [0, 1]$ we set $\mu_t := (\mathbf{i} + t(\mathbf{t}_\mu^\nu - \mathbf{i}))_\# \mu$ and we recall that the λ -convexity yields

$$\frac{\phi(\mu_t) - \phi(\mu)}{t} \leq \phi(\nu) - \phi(\mu) - \frac{\lambda}{2} (1-t) W_2^2(\mu, \nu). \quad (4.41)$$

On the other hand, since $W_2(\mu, \mu_t) = t W_2(\mu, \nu)$, Fréchet differentiability yields

$$\begin{aligned} \liminf_{t \downarrow 0} \frac{\phi(\mu_t) - \phi(\mu)}{t} & \geq \liminf_{t \rightarrow 0^+} \frac{1}{t} \int_{\mathbb{R}^d} \langle \xi(x), \mathbf{t}_\mu^{\mu_t}(x) - x \rangle d\mu(x) \\ & \geq \int_{\mathbb{R}^d} \langle \xi(x), \mathbf{t}_\mu^\nu(x) - x \rangle d\mu(x), \end{aligned}$$

since $\mathbf{t}_\mu^{\mu_t}(x) = x + t(\mathbf{t}_\mu^v(x) - x)$.

In the case of the slope (4.39), (4.41) and the fact that

$$\liminf_{t \downarrow 0} \frac{\phi(\mu_t) - \phi(\mu)}{t} \geq -|\partial\phi|(\mu) W_2(\mu, v) \quad (4.42)$$

yield (4.39). $\square$

A simple consequence of (4.40) is the lower semicontinuity of the slope:

$$\mu_n \rightarrow \mu \quad \text{in } \mathcal{P}_2(\mathbb{R}^d) \implies \liminf_{n \rightarrow \infty} |\partial\phi|(\mu_n) \geq |\partial\phi|(\mu). \quad (4.43)$$

Indeed, if $v \neq \mu$ then $v \neq \mu_n$ for n large enough, hence

$$\liminf_{n \rightarrow \infty} \frac{\phi(\mu_n) - \phi(v)}{W_2(\mu_n, v)} + \lambda W_2(\mu_n, v) \geq \frac{\phi(\mu) - \phi(v)}{W_2(\mu, v)} + \lambda W_2(\mu, v).$$

By estimating the left-hand side with $\liminf_n |\partial\phi|(\mu_n)$ and taking the supremum w.r.t. v , we obtain (4.43).

C. Convexity and strong-weak closure. The next step is to show the closure of the graph of $\partial\phi$: here one has to be careful in the meaning of the convergence of vectors $\xi_n \in L^2(\mu_n; \mathbb{R}^m)$, which belongs to different L^2 -spaces, and we will adopt the following natural one.

DEFINITION 4.5. Let $(\mu_n) \subset \mathcal{P}(\mathbb{R}^d)$ be narrowly converging to μ in $\mathcal{P}(\mathbb{R}^d)$ and let $\mathbf{v}_n \in L^1(\mu_n; \mathbb{R}^m)$. We say that $\mathbf{v}_n$ weakly converge to $\mathbf{v} \in L^1(\mu; \mathbb{R}^m)$ if

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \zeta(x) \mathbf{v}_n(x) d\mu_n(x) = \int_{\mathbb{R}^d} \zeta(x) \mathbf{v}(x) d\mu(x) \quad \forall \zeta \in C_c^\infty(\mathbb{R}^d). \quad (4.44)$$

Clearly, if $\|\mathbf{v}_n\|_{L^1(\mu_n; \mathbb{R}^m)}$ is bounded, a density argument shows that the convergence above is equivalent to the narrow convergence (i.e., in the duality with $C_b(\mathbb{R}^d)$) of the vector-valued measures $\mathbf{v}_n \mu_n$ to $\mathbf{v} \mu$. We now state (see [9], Theorem 5.4.4, for a more general statement) some basic properties of this convergence.

THEOREM 4.6. Let $(\mu_n) \subset \mathcal{P}_2(\mathbb{R}^d)$ be converging to μ in $\mathcal{P}_2(\mathbb{R}^d)$ and let $\mathbf{v}_n \in L^2(\mu_n; \mathbb{R}^m)$ be such that

$$\sup_{n \in \mathbb{N}} \int_{\mathbb{R}^d} |\mathbf{v}_n(x)|^2 d\mu_n(x) < +\infty. \quad (4.45)$$

Then the sequence $(\mathbf{v}_n)$ has weak limit points as $n \rightarrow \infty$, and if $\mathbf{v}$ is any limit point, along

some subsequence $n(k)$, we have

$$\int_{\mathbb{R}^d} |\mathbf{v}(x)|^2 d\mu(x) \leq \liminf_{k \rightarrow \infty} \int_{\mathbb{R}^d} |\mathbf{v}_{n(k)}(x)|^2 d\mu_{n(k)}, \quad (4.46)$$

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} \langle \mathbf{v}_{n(k)}, \varphi \rangle d\mu_{n(k)}(x) = \int_{\mathbb{R}^d} \langle \mathbf{v}(x), \varphi \rangle d\mu(x), \quad (4.47)$$

for every continuous function $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^m$ with at most linear growth.

PROOF. The first statement is a direct consequence of the lower semicontinuity of the relative entropy functional (3.36), in the case when $F(z) = z^2$, see Lemma 3.17 (here actually only the narrow convergence of the μ_n is needed). The convergence property (4.47) follows by a simple truncation argument, taking into account that, $|x|^2$ is uniformly integrable w.r.t. $\{\mu_n\}_{n \in \mathbb{N}}$. $\square$

LEMMA 4.7 (Closure of the subdifferential). *Let ϕ be a λ -convex functional satisfying (4.31a), let (μ_n) be converging to $\mu \in D(\phi)$ in $\mathcal{P}_2(\mathbb{R}^d)$, let $\xi_n \in \partial\phi(\mu_n)$ be satisfying*

$$\sup_n \int_{\mathbb{R}^d} |\xi_n(x)|^2 d\mu_n(x) < +\infty, \quad (4.48)$$

and converging to ξ according to Definition 4.5. Then $\xi \in \partial\phi(\mu)$.

PROOF. Let $v \in D(\phi)$ and let C be the constant in (4.48). We have to pass to the limit as $n \rightarrow \infty$ in the subdifferential inequality

$$\phi(v) - \phi(\mu_n) \geq \int_{\mathbb{R}^d} \langle \xi_n(x), \mathbf{t}_{\mu_n}^v(x) - x \rangle d\mu_n(x) + \frac{\lambda}{2} W_2^2(\mu_n, v). \quad (4.49)$$

By the lower semicontinuity of ϕ the upper limit of $\phi(v) - \phi(\mu_n)$ is less than $\phi(v) - \phi(\mu)$. Passing to the right-hand side, given $\varepsilon > 0$ we choose $\bar{\mathbf{t}} \in C_b^0(\mathbb{R}^d; \mathbb{R}^d)$ such that $\|\mathbf{t}_{\mu}^v - \bar{\mathbf{t}}\|_{L^2(\mu; \mathbb{R}^d)} < \varepsilon^2$ and split the integrals as

$$\int_{\mathbb{R}^d} \langle \xi_n(x), \mathbf{t}_{\mu_n}^v(x) - \bar{\mathbf{t}}(x) \rangle d\mu_n(x) + \int_{\mathbb{R}^d} \langle \xi_n(x), \bar{\mathbf{t}}(x) - x \rangle d\mu_n(x). \quad (4.50)$$

By the Young inequality, the first integrals can be estimated with

$$\frac{C\varepsilon}{2} + \frac{1}{2\varepsilon} \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d} |\mathbf{t}_{\mu_n}^v - \bar{\mathbf{t}}|^2 d\mu_n = \frac{C\varepsilon}{2} + \frac{1}{2\varepsilon} \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - \bar{\mathbf{t}}(x)|^2 d\gamma_n,$$

where $\gamma_n = (\mathbf{i} \times \mathbf{t}_{\mu_n}^v)_{\#} \mu_n$ are the optimal plans induced by $\mathbf{t}_{\mu_n}^v$. Now, by Proposition 7.1.3 of [9] (showing that optimal plans are stable under narrow convergence), we know that γ_n narrowly converge to the plan $\gamma = (\mathbf{i} \times \mathbf{t}_{\mu}^v)_{\#} \mu$ induced by $\mathbf{t}_{\mu}^v$; moreover, as $|y|^2$ is

uniformly integrable with respect to $\{\gamma_n\}$ (because the second marginal of γ_n is constant), Lemma 1.2 gives that the upper limits above are less than

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |y - \bar{\mathbf{t}}(x)|^2 d\gamma = \int_{\mathbb{R}^d \times \mathbb{R}^d} |\mathbf{t}_\mu^v - \bar{\mathbf{t}}|^2 d\mu \leq \varepsilon^2.$$

Summing up, we proved that the limsup of the first integrals in (4.50) is less than $(C + 1)\varepsilon/2$. The convergence of the second integrals in (4.50) to

$$\int_{\mathbb{R}^d} \langle \xi(x), \bar{\mathbf{t}}(x) - x \rangle d\mu(x)$$

follows directly from (4.47) of Theorem 4.6. As a consequence

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d} \langle \xi_n(x), \mathbf{t}_{\mu_n}^v(x) - x \rangle d\mu_n(x) \\ & \geq \int_{\mathbb{R}^d} \langle \xi(x), \mathbf{t}_\mu^v(x) - x \rangle d\mu(x) - \frac{\varepsilon}{2}(C + 1) - \int_{\mathbb{R}^d} |\xi(x)| \cdot |\bar{\mathbf{t}}(x) - \mathbf{t}_\mu^v| d\mu(x). \end{aligned}$$

As ε is arbitrary, the variational inequality (4.49) passes to the limit. $\square$

4.4. Regular functionals

DEFINITION 4.8. A functional $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ satisfying (4.31a) is *regular* if, whenever the *strong* subdifferentials $\xi_n \in \partial\phi(\mu_n)$, $\varphi_n = \phi(\mu_n)$ satisfy

$$\begin{cases} \mu_n \rightarrow \mu & \text{in } \mathcal{P}_2(\mathbb{R}^d), \\ \xi_n \rightarrow \xi & \text{weakly, according to Definition 4.5,} \end{cases} \quad \varphi_n \rightarrow \varphi, \quad \sup_n \|\xi_n\|_{L^2(\mu_n; \mathbb{R}^d)} < +\infty \quad (4.51)$$

then $\xi \in \partial\phi(\mu)$ and $\varphi = \phi(\mu)$.

We just proved that λ -convex functionals are indeed regular.

In the “differential” proof of the convergence of the implicit Euler scheme for gradient flows we will use the following time-dependent variant of Lemma 4.7 whose proof uses the same approximation arguments.

REMARK 4.9. Let $\mu_t^n : [0, T] \rightarrow \mathcal{P}_2^a(\mathbb{R}^d)$ be uniformly bounded and pointwise converging in $[0, T]$ to $\mu_t : [0, T] \rightarrow \mathcal{P}_2^a(\mathbb{R}^d)$ as $n \rightarrow \infty$. Let $\xi^n, \xi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be such that

$$\sup_n \int_0^T \int_{\mathbb{R}^d} |\xi^n|^2 d\mu_t^n dt < +\infty$$

and

$$\lim_{n \rightarrow \infty} \int_0^T \int_{\mathbb{R}^d} \xi^n \varphi d\mu_t^n dt = \int_0^T \int_{\mathbb{R}^d} \xi \varphi d\mu_t dt \quad \forall \varphi \in C_c^\infty([0, T] \times \mathbb{R}^d).$$

Then, for all $v \in \mathcal{P}_2(\mathbb{R}^d)$, we have

$$\lim_{n \rightarrow \infty} \int_0^T \int_{\mathbb{R}^d} \langle \mathbf{t}_{\mu_t^n}^v - \mathbf{i}, \xi^n \rangle d\mu_t^n dt = \int_0^T \int_{\mathbb{R}^d} \langle \mathbf{t}_{\mu_t}^v - \mathbf{i}, \xi \rangle d\mu_t dt.$$

D. Minimal selection and slope.

LEMMA 4.10. *Let ϕ be a regular functional satisfying (4.31a,b). $\mu \in D(|\partial\phi|)$ if and only if $\partial\phi(\mu)$ is not empty and*

$$|\partial\phi|(\mu) = \min \{ \|\xi\|_{L^2(\mu; \mathbb{R}^d)} : \xi \in \partial\phi(\mu) \}, \quad (4.52)$$

where the metric slope $|\partial\phi|(\mu)$ is defined in (4.4).

By the convexity of $\partial\phi(\mu)$ there exists a unique vector $\xi \in \partial\phi(\mu)$ which attains the minimum in (4.52): we will denote it by $\partial^\circ\phi(\mu)$, it belongs to $\text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ and it is also a strong subdifferential.

PROOF. It is clear from the very definition of Fréchet subdifferential that

$$|\partial\phi|(\mu) \leq \|\xi\|_{L^2(\mu; \mathbb{R}^d)} \quad \forall \xi \in \partial\phi(\mu);$$

thus we should prove that if $|\partial\phi|(\mu) < +\infty$ there exists $\xi \in \partial\phi(\mu)$ such that $\|\xi\|_{L^2(\mu; \mathbb{R}^d)} \leq |\partial\phi|(\mu)$. We argue by approximation: for $\mu \in D(|\partial\phi|)$ and $\tau \in (0, \tau_*)$, let μ_τ be a minimizer of (4.31b); by Lemma 4.4 we know that

$$\xi_\tau = \frac{1}{\tau} (\mathbf{t}_{\mu_\tau}^\mu - \mathbf{i}) \in \partial\phi(\mu_\tau), \quad \int_{\mathbb{R}^d} |\xi_\tau(x)|^2 d\mu_\tau(x) = \frac{W_2^2(\mu, \mu_\tau)}{\tau^2},$$

and ξ_τ is a strong subdifferential. Furthermore, it is proved in Lemma 3.1.5 of [9] (in a general metric space setting) that there exists a sequence $(\tau_n) \downarrow 0$ such that

$$\lim_{n \rightarrow \infty} \frac{W_2^2(\mu_{\tau_n}, \mu)}{\tau_n^2} = |\partial\phi|^2(\mu). \quad (4.53)$$

By Theorem 4.6 we know that ξ_τ has some limit point $\xi \in L^2(\mu; \mathbb{R}^d)$ as $\tau \downarrow 0$, according to Definition 4.5. By (4.51) we get $\xi \in \partial\phi(\mu)$ with $\|\xi\|_{L^2(\mu; \mathbb{R}^d)} \leq |\partial\phi|(\mu)$, so that ξ is the (unique) element of minimal norm in $\partial\phi(\mu)$.

By (4.22) we also deduce that $\xi \in \text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$ and Proposition 4.2 shows that ξ is a strong subdifferential. $\square$

REMARK 4.11 (The λ -convex case). When ϕ satisfies the λ -convexity assumption (4.36), the proof of property (4.53) is considerably easier, since μ_τ satisfies the a priori bound ([9], Theorem 3.1.6)

$$(1 + \lambda\tau) \frac{W_2(\mu_\tau, \mu)}{\tau} \leq |\partial\phi|(\mu). \quad (4.54)$$

Indeed, we choose $\mu_t := (\mathbf{i} + t(\mathbf{t}_\mu^{\mu_\tau} - \mathbf{i}))_\# \mu$ and we recall that λ -convexity of ϕ yields

$$\begin{aligned} & \frac{1}{2\tau} W_2^2(\mu, \mu_\tau) + \phi(\mu_\tau) \\ & \leq \frac{1}{2\tau} W_2^2(\mu, \mu_t) + \phi(\mu_t) \\ & \leq \frac{t}{2\tau} (t - \lambda\tau(1-t)) W_2^2(\mu, \mu_\tau) + (1-t)\phi(\mu) + t\phi(\mu_\tau). \end{aligned}$$

Since the right-hand quadratic function has a minimum for $t = 1$, taking the left derivative we obtain

$$\left(\frac{\lambda}{2} + \frac{1}{\tau} \right) W_2^2(\mu, \mu_\tau) + \phi(\mu_\tau) - \phi(\mu) \leq 0,$$

and therefore, by (4.40)

$$\begin{aligned} \frac{1}{2}(1 + \lambda\tau) \frac{W_2^2(\mu, \mu_\tau)}{\tau^2} & \leq \frac{\phi(\mu) - \phi(\mu_\tau)}{\tau} - \frac{W_2^2(\mu, \mu_\tau)}{2\tau^2} \\ & \leq |\partial\phi|(\mu) \frac{W_2(\mu_\tau, \mu)}{\tau} - (1 + \lambda\tau) \frac{W_2^2(\mu_\tau, \mu)}{2\tau^2} \\ & \leq \frac{1}{2(1 + \lambda\tau)} |\partial\phi|^2(\mu), \end{aligned}$$

which yields (4.54).

E. Chain rule. Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ be a regular functional satisfying (4.31a), (4.31b), and let $\mu : (a, b) \mapsto \mu_t \in D(\phi) \subset \mathcal{P}_2(\mathbb{R}^d)$ be an absolutely continuous curve with tangent velocity vector $\mathbf{v}_t$. Let $\Lambda \subset (a, b)$ be the set of points $t \in (a, b)$ such that

- (a) $|\partial\phi|(\mu_t) < +\infty$;
- (b) $\phi \circ \mu$ is differentiable at t ;
- (c) condition (2.56) of Proposition 2.20 holds.

Then

$$\frac{d}{dt} \phi(\mu_t) = \int_{\mathbb{R}^d} \langle \xi_t(x), \mathbf{v}_t(x) \rangle d\mu_t(x) \quad \forall \xi_t \in \partial\phi(\mu_t), \quad \forall t \in \Lambda. \quad (4.55)$$

Moreover, if ϕ is λ -convex along geodesics and

$$\int_a^b |\partial\phi|(\mu_t) |\mu'| (t) dt < +\infty, \quad (4.56)$$

then the map $t \mapsto \phi(\mu_t)$ is absolutely continuous, and $(a, b) \setminus \Lambda$ is $\mathcal{L}^1$ -negligible.

PROOF. Let $\bar{t} \in \Lambda$; observing that

$$\mathbf{v}_h := \frac{1}{h}(\mathbf{t}_{\mu_{\bar{t}}+h}^{\mu_{\bar{t}}+h} - \mathbf{i}) \rightarrow \mathbf{v}_{\bar{t}} \quad \text{in } L^2(\mu_{\bar{t}}; \mathbb{R}^d), \quad (4.57)$$

we have

$$\phi(\mu_{\bar{t}+h}) - \phi(\mu_{\bar{t}}) \geq h \int_{\mathbb{R}^d} \langle \mathbf{v}_h(x), \boldsymbol{\xi}_{\bar{t}}(x) \rangle d\mu_{\bar{t}}(x) + o(h). \quad (4.58)$$

Dividing by h and taking the right and left limits as $h \rightarrow 0$ we obtain that the left and right derivatives $d/dt_{\pm}\phi(\mu_t)$ satisfy

$$\begin{aligned} \left. \frac{d}{dt_+} \phi(\mu_t) \right|_{t=\bar{t}} &\geq \int_{\mathbb{R}^d} \langle \mathbf{v}_{\bar{t}}(x), \boldsymbol{\xi}_{\bar{t}}(x) \rangle d\mu_{\bar{t}}(x), \\ \left. \frac{d}{dt_-} \phi(\mu_t) \right|_{t=\bar{t}} &\leq \int_{\mathbb{R}^d} \langle \mathbf{v}_{\bar{t}}(x), \boldsymbol{\xi}_{\bar{t}}(x) \rangle d\mu_{\bar{t}}(x) \end{aligned}$$

and therefore we find (4.55).

In the λ -convex case, using (4.40) it can be shown (see Corollary 2.4.10 in [9]) that (4.56) implies that $t \mapsto \phi(\mu_t)$ is absolutely continuous in (a) and (b) and thus conditions (a)–(c) hold $\mathcal{L}^1$ -a.e. in (a) and (b). $\square$

4.5. Examples of subdifferentials

In this section we consider in the detail the subdifferential of the convex functionals presented in Section 3.2 (potential energy, interaction energy, internal energy, negative Wasserstein distance), with a particular attention to the characterization of the elements with minimal norm.

We start by considering a general, but *smooth*, situation.

4.5.1. Variational integrals: the smooth case. In order to clarify the underlying structure of many examples and the link between the notion of Wasserstein subdifferential and the standard variational calculus for integral functionals, we first consider the case of a variational integral of the type

$$\mathcal{F}(\mu) := \begin{cases} \int_{\mathbb{R}^d} F(x, u(x), \nabla u(x)) dx & \text{if } \mu = u \cdot \mathcal{L}^d \text{ with } u \in C^1(\mathbb{R}^d), \\ +\infty & \text{otherwise.} \end{cases} \quad (4.59)$$

Since we are not claiming any generality and we are only interested in the form of the subdifferential, we will assume enough regularity to justify all the computations; therefore, we suppose that $F: \mathbb{R}^d \times [0, +\infty) \times \mathbb{R}^d \rightarrow [0, +\infty)$ is a C^2 function with $F(x, 0, p) = 0$ for every $x, p \in \mathbb{R}^d$ and we consider the case of a smooth and strictly positive density u :

as usual, we denote by $(x, z, p) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$ the variables of F and by $\delta\mathcal{F}/\delta u$ the first variation density

$$\frac{\delta\mathcal{F}}{\delta u}(x) := F_z(x, u(x), \nabla u(x)) - \nabla \cdot F_p(x, u(x), \nabla u(x)). \quad (4.60)$$

LEMMA 4.12. *If $\mu = u \cdot \mathcal{L}^d \in \mathcal{P}_2^a(\mathbb{R}^d)$ with $u \in C^2(\mathbb{R}^d)$ satisfies $\mathcal{F}(\mu) < +\infty$ and $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$ belongs to the strong subdifferential of $\mathcal{F}$ at μ (in particular, by Proposition 4.2, if $\mathbf{w} \in \partial\phi(\mu) \cap \text{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)$), then*

$$\mathbf{w}(x) = \nabla \frac{\delta\mathcal{F}}{\delta u}(x) \quad \text{for } \mu\text{-a.e. } x \in \mathbb{R}^d, \quad (4.61)$$

and for every vector field $\xi \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d} \langle \mathbf{w}(x), \xi(x) \rangle d\mu(x) = - \int_{\mathbb{R}^d} \frac{\delta\mathcal{F}}{\delta u}(x) \nabla \cdot (u(x)\xi(x)) dx. \quad (4.62)$$

PROOF. We take a smooth vector field $\xi \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$ and we set for $\varepsilon \in \mathbb{R}$ sufficiently small $\mu_\varepsilon := (\mathbf{i} + \varepsilon\xi)_\# \mu$. If $\mathbf{w}$ is a strong subdifferential, we know that

$$\begin{aligned} \limsup_{\varepsilon \uparrow 0} \frac{\mathcal{F}(\mu_\varepsilon) - \mathcal{F}(\mu)}{\varepsilon} &\leq \int_{\mathbb{R}^d} \langle \mathbf{w}(x), \xi(x) \rangle d\mu(x) \\ &\leq \liminf_{\varepsilon \downarrow 0} \frac{\mathcal{F}(\mu_\varepsilon) - \mathcal{F}(\mu)}{\varepsilon}; \end{aligned} \quad (4.63)$$

on the other hand, by the change of variables formula we know that $\mu_\varepsilon = u_\varepsilon \mathcal{L}^d$ with

$$u_\varepsilon(y) = \frac{u}{\det(I + \varepsilon \nabla \xi)} \circ (\mathbf{i} + \varepsilon \xi)^{-1}(y) \quad \forall y \in \mathbb{R}^d. \quad (4.64)$$

The map $(x, \varepsilon) \mapsto u_\varepsilon(x)$ is of class C^2 with $u_\varepsilon(x) = u(x)$ outside a compact set and

$$u_\varepsilon(x)|_{\varepsilon=0} = u(x), \quad \left. \frac{\partial u_\varepsilon(x)}{\partial \varepsilon} \right|_{\varepsilon=0} = -\nabla \cdot (u(x)\xi(x)). \quad (4.65)$$

Standard variational formulae (see, e.g., [53], Vol. I, Section 1.2.1) yield

$$\lim_{\varepsilon \rightarrow 0} \frac{\mathcal{F}(\mu_\varepsilon) - \mathcal{F}(\mu)}{\varepsilon} = - \int_{\mathbb{R}^d} \frac{\delta\mathcal{F}}{\delta u}(x) \nabla \cdot (u(x)\xi(x)) dx, \quad (4.66)$$

which shows (4.62). □

4.5.2. The potential energy. Let $V: \mathbb{R}^d \rightarrow (-\infty, +\infty]$ be a proper, l.s.c. and λ -convex functional and let $\mathcal{V}(\mu) = \int_{\mathbb{R}^d} V d\mu$ be defined on $\mathcal{P}_2(\mathbb{R}^d)$. We denote by $\text{graph } \partial V$ the

graph of the Fréchet subdifferential of V in $\mathbb{R}^d \times \mathbb{R}^d$, i.e., the subset of the couples $(x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d$ satisfying

$$V(x_3) \geq V(x_1) + \langle x_2, x_3 - x_1 \rangle + \frac{\lambda}{2} |x_1 - x_2|^2 \quad \forall x_3 \in \mathbb{R}^d. \quad (4.67)$$

As usual, $\partial^\circ V(x)$ denotes the element of minimal norm in $\partial V(x)$.

Notice that the potential energy functional (as well as the interaction energy functional) fails to satisfy (4.31a), and for this reason it would be more appropriate to consider a more general notion of subdifferential, involving plans and not only maps as elements of the subdifferential, and, at the same time, taking into account transport plans and not only transport maps (see Section 10.3 of [9]).

In the present case, we choose an intermediate generalization, and say that $\xi \in L^2(\mu; \mathbb{R}^d)$ belongs to the Fréchet subdifferential $\partial \mathcal{V}(\mu)$ at $\mu \in D(\mathcal{V})$ if

$$\mathcal{V}(v) - \mathcal{V}(\mu) \geq \inf_{\gamma \in \Gamma_o(\mu, v)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \xi(x), y - x \rangle d\gamma(x) + o(W_2(\mu, v)). \quad (4.68)$$

The following characterization of $\partial \mathcal{V}$ and of its minimal selection is proved in Proposition 10.4.2 of [9].

PROPOSITION 4.13. *Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\xi \in L^2(\mu; \mathbb{R}^d)$. Then*

- (i) *ξ is a strong subdifferential of $\mathcal{V}$ at μ iff $\xi(x) \in \partial V(x)$ for μ -a.e. x ,*
- (ii) *$\partial^\circ \mathcal{V}(\mu) = \partial^\circ V(x)$ for μ -a.e. $x \in \mathbb{R}^d$.*

4.5.3. The internal energy. Let $\mathcal{F}$ be the functional

$$\mathcal{F}(\mu) := \begin{cases} \int_{\mathbb{R}^d} F(u(x)) d\mathcal{L}^d(x) & \text{if } \mu = u \cdot \mathcal{L}^d \in \mathcal{P}_2^a(\mathbb{R}^d), \\ +\infty & \text{otherwise,} \end{cases} \quad (4.69)$$

for a convex differentiable function satisfying

$$F(0) = 0, \quad \liminf_{s \downarrow 0} \frac{F(s)}{s^\alpha} > -\infty \quad \text{for some } \alpha > \frac{d}{d+2} \quad (4.70)$$

as in Example 3.8. Recall that if F is nonnegative and has superlinear growth at infinity then the functional $\mathcal{F}$ is l.s.c. with respect to the narrow convergence (indeed, under this growth condition the lower semicontinuity can be checked w.r.t. to the stronger weak L^1 convergence, by Dunford–Pettis theorem, and lower semicontinuity w.r.t. weak L^1 convergence is a direct consequence of the convexity of $\mathcal{F}$).

We confine our discussion to the case when F has a more than linear growth at infinity, i.e.,

$$\lim_{z \rightarrow +\infty} \frac{F(z)}{z} = +\infty, \quad (4.71)$$

see Theorems 10.4.6 and 10.4.8 of [9] for a discussion of the (sub)linear case.

We set $L_F(z) = zF'(z) - F(z) : [0, +\infty) \rightarrow [0, +\infty)$ and we observe that L_F is strictly related to the *convex* function

$$G(z, s) := sF\left(\frac{z}{s}\right), \quad z \in [0, +\infty), s \in (0, +\infty), \quad (4.72)$$

since

$$\frac{\partial}{\partial s} G(z, s) = -\frac{z}{s} F'\left(\frac{z}{s}\right) + F\left(\frac{z}{s}\right) = -L_F\left(\frac{z}{s}\right). \quad (4.73)$$

In particular (recall that $F(0) = 0$, by (4.70))

$$G(z, s) \leq F(z) \quad \text{for } s \geq 1, \quad \frac{F(z) - G(z, s)}{s - 1} \uparrow L_F(z) \quad \text{as } s \downarrow 1. \quad (4.74)$$

We will also suppose that F satisfies the condition

$$\text{the map } s \mapsto s^d F(s^{-d}) \text{ is convex and nonincreasing in } (0, +\infty), \quad (4.75)$$

yielding the geodesic convexity of $\mathcal{F}$.

The following lemma shows the existence of the directional derivative of $\mathcal{F}$ along a suitable class of directions including all optimal transport maps.

LEMMA 4.14 (Directional derivative of $\mathcal{F}$). *Suppose that $F : [0, +\infty) \rightarrow \mathbb{R}$ is a convex differentiable function satisfying (4.70), (4.71) and (4.75). Let $\mu = u\mathcal{L}^d \in D(\mathcal{F})$, $\mathbf{r} \in L^2(\mu; \mathbb{R}^d)$ and $\bar{t} > 0$ be such that*

(i) *$\mathbf{r}$ is differentiable $u\mathcal{L}^d$ -a.e. and $\mathbf{r}_t := (1 - t)\mathbf{i} + t\mathbf{r}$ is $u\mathcal{L}^d$ -injective with $|\det \nabla \mathbf{r}_t(x)| > 0$ $u\mathcal{L}^d$ -a.e., for any $t \in [0, \bar{t}]$;*

(ii) *$\nabla \mathbf{r}_{\bar{t}}$ is diagonalizable with positive eigenvalues;*

(iii) *$\mathcal{F}((\mathbf{r}_{\bar{t}})_{\#}\mu) < +\infty$.*

Then the map $t \mapsto t^{-1}(\mathcal{F}((\mathbf{r}_t)_{\#}\mu) - \mathcal{F}^(\mu))$ is nondecreasing in $[0, \bar{t}]$ and*

$$+\infty > \lim_{t \downarrow 0} \frac{\mathcal{F}((\mathbf{r}_t)_{\#}\mu) - \mathcal{F}(\mu)}{t} = - \int_{\mathbb{R}^d} L_F(u) \operatorname{tr} \nabla(\mathbf{r} - \mathbf{i}) \, dx. \quad (4.76)$$

The identity above still holds when assumption (ii) on $\mathbf{r}$ is replaced by

(ii') *$\|\nabla(\mathbf{r} - \mathbf{i})\|_{L^\infty(u\mathcal{L}^d; \mathbb{R}^{d \times d})} < +\infty$ (in particular, if $\mathbf{r} - \mathbf{i} \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$), and F satisfies in addition the “doubling” condition*

$$\exists C > 0: \quad F(z + w) \leq C(1 + F(z) + F(w)) \quad \forall z, w. \quad (4.77)$$

PROOF. By assumptions (i) and (ii), taking into account Lemma 1.3 we have

$$\begin{aligned} \mathcal{F}((\mathbf{r}_t)_{\#}\mu) - \mathcal{F}(\mu) &= \int_{\mathbb{R}^d} F\left(\frac{u(x)}{\det \nabla \mathbf{r}_t(x)}\right) \det \nabla \mathbf{r}_t(x) \, dx - \int_{\mathbb{R}^d} F(u(x)) \, dx \\ &= \int_{\mathbb{R}^d} (G(u(x), \det \nabla \mathbf{r}_t(x)) - F(u(x))) \, dx \end{aligned}$$

for any $t \in (0, \bar{t}]$. Assumption (4.75), together with the concavity of the map $t \mapsto [\det((1 - t)I + t\nabla \mathbf{r})]^{1/d}$, implies that the function

$$\frac{G(u(x), \det \nabla \mathbf{r}_t) - F(u(x))}{t}, \quad t \in (0, \bar{t}], \quad (4.78)$$

is nondecreasing w.r.t. t and bounded above by an integrable function (take $t = \bar{t}$ and apply (iii)). Therefore the monotone convergence theorem gives

$$\lim_{t \downarrow 0} \frac{\mathcal{F}((\mathbf{r}_t)_\# \mu) - \mathcal{F}(\mu)}{t} = \int_{\mathbb{R}^d} \frac{d}{dt} G(u(x), \det \nabla \mathbf{r}_t(x)) \Big|_{t=0} dx$$

and the expansion $\det \nabla \mathbf{r}_t = 1 + t \operatorname{tr}(\nabla(\mathbf{r} - \mathbf{i})) + o(t)$ together with (4.73) give the result.

In the case when (ii') holds, the argument is analogous but, since condition (ii) fails, we cannot rely anymore on the monotonicity of the function in (4.78). However, using the inequalities

$$F(w) - F(0) \leq w F'(w) \leq F(2w) - F(w)$$

and the doubling condition we easily see that the derivative w.r.t. s of the function $G(z, s)$ can be bounded by $C(1 + F^+(z))$ for $|s - 1| \leq 1/2$. Therefore we can use the dominated convergence theorem instead of the monotone convergence theorem to pass to the limit. $\square$

The next technical lemma shows that we can “integrate by parts” in (4.76) preserving the inequality, if $L_F(u)$ is locally in $W^{1,1}$.

LEMMA 4.15 (A “weak” integration by parts formula). *Under the same assumptions of Lemma 4.14, let us suppose that*

- (i) $\operatorname{supp} \mu \subset \bar{\Omega}$, Ω being a convex open subset of $\mathbb{R}^d$ (not necessarily bounded);
- (ii) $L_F(u) \in W_{\operatorname{loc}}^{1,1}(\Omega)$;
- (iii) $K = \operatorname{supp}((\mathbf{r}_{\bar{t}})_\# \mu)$ is a compact subset of Ω for some $\bar{t} \in [0, 1]$;
- (iv) $\mathbf{r} \in BV_{\operatorname{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ and $D \cdot \mathbf{r} \geq 0$.

Then we can find an increasing family of nonnegative Lipschitz functions $\chi_k : \mathbb{R}^d \rightarrow [0, 1]$ with compact support in Ω such that $\chi_k \uparrow \chi_\Omega$ and

$$- \int_{\mathbb{R}^d} L_F(u(x)) \operatorname{tr} \nabla(\mathbf{r} - \mathbf{i}) dx \geq \limsup_{k \rightarrow \infty} \int_{\mathbb{R}^d} \langle \nabla L_F(u), \mathbf{r} - \mathbf{i} \rangle \chi_k dx. \quad (4.79)$$

PROOF. Possibly replacing $\mathbf{r}$ by $\mathbf{r}_{\bar{t}}$, we can assume that $\bar{t} = 1$ in (iii). Let us first recall that by Calderon–Zygmund theorem (see, for instance, [8]) the pointwise divergence $\operatorname{tr}(\nabla \mathbf{r})$ is the absolutely continuous part of the distributional divergence $D \cdot \mathbf{r}$; therefore we have

$$\int_{\mathbb{R}^d} v \operatorname{tr}(\nabla \mathbf{r}) dx \leq - \int_{\mathbb{R}^d} \langle \nabla v, \mathbf{r} \rangle dx, \quad (4.80)$$

provided $v \in C_c^\infty(\mathbb{R}^d)$ is nonnegative. As $\mathbf{r}$ is bounded, by approximation the same inequality remains true for every nonnegative function $v \in W^{1,1}(\mathbb{R}^d)$. For every Lipschitz function $\eta: \mathbb{R}^d \rightarrow [0, 1]$ with compact support in Ω , choosing $v := \eta L_F(u) \in W^{1,1}(\mathbb{R}^d)$ we get

$$\int_{\mathbb{R}^d} (\eta L_F(u)) \operatorname{tr}(\nabla \mathbf{r}) \, dx \leq - \int_{\mathbb{R}^d} \langle \nabla (\eta L_F(u)), \mathbf{r} \rangle \, dx. \quad (4.81)$$

On the other hand, a standard integration by parts yields

$$\int_{\Omega} (\eta L_F(u)) \operatorname{tr}(\nabla \mathbf{i}) \, dx = - \int_{\Omega} \langle \nabla (\eta L_F(u)), \mathbf{i} \rangle \, dx; \quad (4.82)$$

summing up with (4.81) and inverting the sign we find

$$- \int_{\mathbb{R}^d} (\eta L_F(u)) \operatorname{tr}(\nabla (\mathbf{r} - \mathbf{i})) \, dx \geq \int_{\mathbb{R}^d} \langle \nabla (\eta L_F(u)), \mathbf{r} - \mathbf{i} \rangle \, dx. \quad (4.83)$$

Now we choose carefully the test function η . We consider an increasing family bounded open convex sets Ω_k such that

$$\overline{\Omega}_k \subset \subset \Omega, \quad \Omega = \bigcup_{k=1}^{\infty} \Omega_k$$

and for each convex set Ω_k we consider the function

$$\chi_k(x) := kd(x, \mathbb{R}^d \setminus \Omega_k) \wedge 1. \quad (4.84)$$

χ_k is an increasing family of nonnegative Lipschitz functions which take their values in $[0, 1]$ and satisfy $\chi_k(x) \equiv 1$ if $d(x, \mathbb{R}^d \setminus \Omega_k) \geq 1/k$; in particular, $\chi_k \equiv 1$ in K for k sufficiently large. Moreover χ_k is concave in Ω_k , since the distance function $d(\cdot, \mathbb{R}^d \setminus \Omega_k)$ is concave. Choosing $\eta := \chi_k$ in (4.83) we get

$$\begin{aligned} & - \int_{\mathbb{R}^d} (\chi_k L_F(u)) \operatorname{tr}(\nabla (\mathbf{r} - \mathbf{i})) \, dx \\ & \geq \int_{\mathbb{R}^d} \langle \nabla L_F(u), \mathbf{r} - \mathbf{i} \rangle \chi_k \, dx + \int_{\Omega_k} \langle \nabla \chi_k, \mathbf{r} - \mathbf{i} \rangle L_F(u) \, dx \\ & \geq \int_{\mathbb{R}^d} \langle \nabla L_F(u), \mathbf{r} - \mathbf{i} \rangle \chi_k \, dx \end{aligned} \quad (4.85)$$

since the second integrand of (4.85) is nonnegative: in fact, for $\mathcal{L}^d$ -a.e. $x \in \Omega_k$ where $L_F(u(x))$ is strictly positive, the concavity of χ_k and $\mathbf{r}(x) \in K$ yields

$$\langle \nabla \chi_k(x), \mathbf{r}(x) - \mathbf{i}(x) \rangle \geq \chi_k(\mathbf{r}(x)) - \chi_k(x) = 1 - \chi_k(x) \geq 0.$$

Passing to the limit as $k \rightarrow \infty$ in the previous integral inequality, we obtain (4.79) (recall that the function in the left-hand side of (4.79) is semiintegrable by (4.76)). $\square$

In the following theorem we characterize the minimal selection in the subdifferential of $\mathcal{F}$ and give, under the doubling condition, a formula for the slope of the functional.

THEOREM 4.16 (Slope and subdifferential of $\mathcal{F}$). *Let $F : [0, +\infty) \rightarrow \mathbb{R}$ be a convex differentiable function satisfying (4.70), (4.71), (4.75) and (4.77). Assume that $\mathcal{F}$ has finite slope at $\mu = u\mathcal{L}^d \in \mathcal{P}_2^a(\mathbb{R}^d)$. Then $L_F(u) \in W^{1,1}(\mathbb{R}^d)$, $\nabla L_F(u) = \mathbf{w}u$ for some function $\mathbf{w} \in L^2(u\mathcal{L}^d; \mathbb{R}^d)$ and*

$$\left(\int_{\mathbb{R}^d} |\mathbf{w}(x)|^2 u(x) dx \right)^{1/2} = |\partial\mathcal{F}|(\mu) < +\infty. \quad (4.86)$$

Conversely, if $\nabla L_F(u) \in W_{\text{loc}}^{1,1}(\mathbb{R}^d)$ and $\nabla L_F(u) = \mathbf{w}u$ for some $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$, then $\mathcal{F}$ has a finite slope at $\mu = u\mathcal{L}^d$ and $\mathbf{w} = \partial^\circ \mathcal{F}(\mu)$.

PROOF. (a) We apply first (4.76) with $\mathbf{r} = 2\mathbf{i}$ and take into account that

$$W_2(\mu, ((1-t)\mathbf{i} + t\mathbf{r})_\# \mu) \leq t \|\mathbf{i}\|_{L^2(u\mathcal{L}^d; \mathbb{R}^d)}$$

to obtain

$$d \int_{\mathbb{R}^d} L_F(u) dx \leq |\partial\mathcal{F}|(\mu) \|\mathbf{i}\|_{L^2(u\mathcal{L}^d; \mathbb{R}^d)},$$

so that $L_F(u) \in L^1(\mathbb{R}^d)$. Next, we apply (4.76) with $\mathbf{r} - \mathbf{i}$ equal to a $C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$ function $\mathbf{t}$ (notice that condition (i) holds with $\bar{t} < \sup |\nabla \mathbf{t}|$) and use again the inequality $W_2(\mu, ((1-t)\mathbf{i} + t\mathbf{r})_\# \mu) \leq t \|\mathbf{r} - \mathbf{i}\|_{L^2(u\mathcal{L}^d)}$ to obtain

$$\int_{\mathbb{R}^d} L_F(u) \operatorname{tr}(\nabla \mathbf{t}) dx \leq |\partial\mathcal{F}^*|(\mu) \|\mathbf{t}\|_{L^2(u\mathcal{L}^d)} \leq |\partial\mathcal{F}^*|(\mu) \sup_{\mathbb{R}^d} |\mathbf{t}|.$$

As $\mathbf{t}$ is arbitrary, Riesz theorem gives that $L_F(u)$ is a function of bounded variation (i.e., its distributional derivative $DL_F(u)$ is a finite $\mathbb{R}^d$ -valued measure in $\mathbb{R}^d$), so that we can rewrite the inequality as

$$\left| \sum_{i=1}^d \int_{\mathbb{R}^d} \mathbf{t}_i dD_i L_F(u) \right| \leq |\partial\mathcal{F}|(\mu) \|\mathbf{t}\|_{L^2(u\mathcal{L}^d; \mathbb{R}^d)}.$$

By L^2 duality theory there exists $\mathbf{w} \in L^2(u\mathcal{L}^d; \mathbb{R}^d)$ with $\|\mathbf{w}\|_2 \leq |\partial\mathcal{F}|(\mu)$ such that

$$\sum_{i=1}^d \int_{\mathbb{R}^d} \mathbf{t}_i dD_i L_F(u) = \int_{\mathbb{R}^d} \langle \mathbf{w}, \mathbf{t} \rangle d(u\mathcal{L}^d) \quad \forall \mathbf{t} \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d).$$

Therefore $L_F(u) \in W^{1,1}(\mathbb{R}^d)$ and $\nabla L_F(u) = \mathbf{w}u$. This leads to the inequality $\leq$ in (4.86).

In order to show that equality holds in (4.86) we will prove that $\mathbf{w}$ belongs to $\partial\mathcal{F}(\mu)$. We have to show that (4.37) holds for any $\nu \in D(\mathcal{F})$. Using the doubling condition it is also easy to find a sequence of measures ν_h with compact support converging to ν in $\mathcal{P}_2(\mathbb{R}^d)$ and such that $\mathcal{F}(\nu_h)$ converges to $\mathcal{F}(\nu)$, hence we can also assume that $\text{supp } \nu$ is compact.

As $\mathbf{t}_\mu^\nu$ is induced by the gradient of a Lipschitz and convex map φ , we know that all the conditions of Lemma 4.14 are fulfilled with $\mathbf{r} = \nabla\varphi$, and also Lemma 4.15 holds; therefore, by applying (4.76), the geodesic convexity of $\mathcal{F}$, and (4.79) we obtain

$$\begin{aligned} \mathcal{F}(\nu) - \mathcal{F}(\mu) &\geq \limsup_{h \rightarrow \infty} \int_{\mathbb{R}^d} \langle \nabla L_F(u), (\mathbf{r} - \mathbf{i}) \rangle \chi_h \, dx \\ &= \limsup_{h \rightarrow \infty} \int_{\mathbb{R}^d} \langle \mathbf{w}, (\mathbf{r} - \mathbf{i}) \rangle \chi_h u \, dx \\ &= \int_{\mathbb{R}^d} \langle \mathbf{w}, \mathbf{r} - \mathbf{i} \rangle \, d\mu, \end{aligned}$$

proving that $\mathbf{w} \in \partial\mathcal{F}(\mu)$.

Finally, we notice that our proof that $\mathbf{w} = \nabla L_F(u)/u \in \partial\mathcal{F}(\mu)$ does not use the finiteness of slope, but only the assumption $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$, therefore these conditions imply that the subdifferential is not empty and that the slope is finite. $\square$

4.5.4. The relative internal energy. In this section we briefly discuss the modifications which should be apported to the previous results, when one consider a relative energy functional as in Section 3.3.

We thus consider a log-concave probability measure $\gamma = e^{-V} \mathcal{L}^d \in \mathcal{P}(\mathbb{R}^d)$ induced by a convex l.s.c. potential

$$V : \mathbb{R}^d \rightarrow (-\infty, +\infty], \quad \text{with } \Omega = \text{int } D(V) \neq \emptyset. \quad (4.87)$$

We are also assuming that the energy density

$$\begin{aligned} F : [0, +\infty) &\rightarrow [0, +\infty] \text{ is convex and l.s.c.,} \\ \text{it satisfies the doubling property (4.77)} \\ \text{and the geodesic convexity condition (3.22),} \end{aligned} \quad (4.88)$$

which yield that the map $s \mapsto \widehat{F}(s) := F(e^{-s})e^s$ is convex and nonincreasing in $\mathbb{R}$. The functional

$$\mathcal{F}(\mu|\gamma) := \int_{\mathbb{R}^d} F(\rho) \, d\gamma = \int_{\Omega} F\left(\frac{u}{e^{-V}}\right) e^{-V} \, dx, \quad \mu = \rho \cdot \gamma = u \mathcal{L}^d \quad (4.89)$$

is therefore geodesically convex in $\mathcal{P}_2(\mathbb{R}^d)$, by Theorem 3.23. It is easy to check that whenever $\widehat{F}$ is not constant (case which corresponds to a linear F and a constant functional $\mathcal{F}$), F has a superlinear growth and therefore $\mathcal{F}$ is lower semicontinuous in $\mathcal{P}_2(\mathbb{R}^d)$.

THEOREM 4.17 (Subdifferential of $\mathcal{F}(\cdot|\gamma)$). *The functional $\mathcal{F}(\cdot|\gamma)$ has finite slope at $\mu = \rho\gamma = u\mathcal{L}^d \in D(\mathcal{F})$ if and only if $L_F(\rho) \in W_{\text{loc}}^{1,1}(\Omega)$ and $\nabla L_F(\rho) = \rho\mathbf{w}$ for some function $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$. In this case*

$$\left(\int_{\mathbb{R}^d} |\mathbf{w}(x)|^2 d\mu(x) \right)^{1/2} = |\partial\mathcal{F}|(\mu), \quad (4.90)$$

and $\mathbf{w} = \partial^\circ \mathcal{F}(\mu)$.

PROOF. We argue as in Theorem 4.16: in the present case the directional derivative formula (4.76) becomes

$$\begin{aligned} +\infty &> \lim_{t \downarrow 0} \frac{\mathcal{F}((\mathbf{r}_t)_\# \mu|\gamma) - \mathcal{F}(\mu|\gamma)}{t} \\ &= - \int_{\mathbb{R}^d} L_F(u/e^{-V}) (e^{-V} \operatorname{tr} \nabla(\mathbf{r} - \mathbf{i}) - e^{-V} \langle \nabla V, \mathbf{r} - \mathbf{i} \rangle) dx \\ &= - \int_{\mathbb{R}^d} L_F(\rho) \operatorname{tr} \nabla(e^{-V}(\mathbf{r} - \mathbf{i})) dx \end{aligned} \quad (4.91)$$

for every vector field $\mathbf{r}$ satisfying the assumptions of Lemma 4.14 and $\mathcal{F}(\mathbf{r}_\# \mu|\gamma)$ is finite. Choosing as before $\mathbf{r} = \mathbf{i} + e^V \mathbf{t}$, $\mathbf{t} \in C_c^\infty(\Omega; \mathbb{R}^d)$, since V is bounded in each compact subset of Ω , we get

$$\int_{\Omega} L_F(\rho) \operatorname{tr} \nabla \mathbf{t} dx \leq |\partial\mathcal{F}|(\mu) \sup_{\mathbb{R}^d} |e^V \mathbf{t}|,$$

so that $L_F(\rho) \in BV_{\text{loc}}(\Omega)$. Choosing now $\mathbf{r} = \mathbf{i} + \mathbf{t}$ with $\mathbf{t} \in C_c^\infty(\Omega; \mathbb{R}^d)$ we get

$$\left| \sum_{i=1}^d \int_{\Omega} \mathbf{t}_i dD_i L_F(\rho) d\gamma \right| \leq |\partial\mathcal{F}|(\mu) \|\mathbf{t}\|_{L^2(\mu; \mathbb{R}^d)}$$

so that there exists $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$ such that

$$\begin{aligned} \sum_{i=1}^d \int_{\Omega} \mathbf{t}_i dD_i L_F(\rho) d\gamma &= \int_{\mathbb{R}^d} \langle \mathbf{w}, \mathbf{t} \rangle d\mu \\ &= \int_{\mathbb{R}^d} \langle u\mathbf{w}, \mathbf{t} \rangle e^{-V} dx \quad \forall \mathbf{t} \in C_c^\infty(\Omega; \mathbb{R}^d), \end{aligned}$$

thus showing that $L_F(\rho) \in W_{\text{loc}}^{1,1}(\Omega)$ and $\nabla L_F(\rho) = u e^{-V} \mathbf{w} = \rho \mathbf{w}$.

Conversely, if $L_F(\rho) \in W_{\text{loc}}^{1,1}(\Omega)$ with $\nabla L_F(\rho) = \rho \mathbf{w}$ and $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$, arguing as in Lemma 4.15 we have for every measure $\nu = \mathbf{r}_\# \mu$ with compact support in Ω

$$\begin{aligned}
 \mathcal{F}(\nu|\gamma) - \mathcal{F}(\mu|\gamma) &\geq \limsup_{k \rightarrow \infty} - \int_{\Omega} L_F(\rho) \operatorname{tr} \nabla(e^{-V}(\mathbf{r} - \mathbf{i})) \chi_k \, d\mathbf{x} \\
 &\geq \limsup_{k \rightarrow \infty} \int_{\Omega} \langle \chi_k \nabla L_F(\rho) + L_F(\rho) \nabla \chi_k, \mathbf{r} - \mathbf{i} \rangle d\gamma \\
 &\geq \limsup_{k \rightarrow \infty} \int_{\Omega} \langle \nabla L_F(\rho), \mathbf{r} - \mathbf{i} \rangle \chi_k \, d\gamma \\
 &\geq \limsup_{k \rightarrow \infty} \int_{\Omega} \langle \mathbf{w}, \mathbf{r} - \mathbf{i} \rangle \chi_k \, d\mu \\
 &= \int_{\Omega} \langle \mathbf{w}, \mathbf{r} - \mathbf{i} \rangle d\mu,
 \end{aligned}$$

which shows, through a density argument, that $\mathbf{w} \in \partial \mathcal{F}(\mu)$. $\square$

4.5.5. The interaction energy. In this section we consider the interaction energy functional $\mathcal{W}: \mathcal{P}_2(\mathbb{R}^d) \rightarrow [0, +\infty]$ defined by

$$\mathcal{W}(\mu) := \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W(x - y) \, d\mu \times \mu(x, y).$$

Without loss of generality we shall assume that $W: \mathbb{R}^d \rightarrow [0, +\infty)$ is an even function; our main assumption, besides the convexity of $\mathbb{R}^d$, is the doubling condition

$$\exists C_W > 0: \quad W(x + y) \leq C_W(1 + W(x) + W(y)) \quad \forall x, y \in \mathbb{R}^d. \quad (4.92)$$

Let us first state a preliminary result: we are denoting by $\bar{\mu}$ the barycenter of the measure μ ,

$$\bar{\mu} := \int_{\mathbb{R}^d} x \, d\mu(x). \quad (4.93)$$

LEMMA 4.18. *Assume that $W: \mathbb{R}^d \rightarrow [0, +\infty)$ is convex, Gateaux differentiable, even, and satisfies the doubling condition (4.92). Then for any $\mu \in D(\mathcal{W})$ we have*

$$\int_{\mathbb{R}^d} W(x) \, d\mu(x) \leq C_W(1 + \mathcal{W}(\mu) + W(\bar{\mu})) < +\infty, \quad (4.94)$$

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |\nabla W(x - y)| \, d\mu \times \mu(x, y) \leq C_W(1 + S_W + \mathcal{W}(\mu)) < +\infty, \quad (4.95)$$

where $S_W := \sup_{|y| \leq 1} W(y)$. In particular $\mathbf{w} := (\nabla W) * \mu$ is well defined for μ -a.e. $x \in \mathbb{R}^d$, it belongs to $L^1(\mu; \mathbb{R}^d)$, and it satisfies

$$\begin{aligned} & \int_{\mathbb{R}^{2d} \times \mathbb{R}^d} \langle \nabla W(x_1 - x_2), y_1 - x_1 \rangle d\boldsymbol{\gamma}(x_1, y_1) d\mu(x_2) \\ &= \int_{\mathbb{R}^{2d}} \langle \mathbf{w}(x_1), y_1 - x_1 \rangle d\boldsymbol{\gamma}(x_1, y_1), \end{aligned} \quad (4.96)$$

for every $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$ with $\nu \in D(W)$. In particular, choosing $\boldsymbol{\gamma} := (\mathbf{i} \times \mathbf{r})_{\#}\mu$, we have

$$\begin{aligned} & \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \nabla W(x - y), \mathbf{r}(x) \rangle d\mu \times \mu(x, y) \\ &= \int_{\mathbb{R}^d} \langle \mathbf{w}(x), \mathbf{r}(x) \rangle d\mu(x) \end{aligned} \quad (4.97)$$

for every vector field $\mathbf{r} \in L^\infty(\mu; \mathbb{R}^d)$ and for $\mathbf{r} := \lambda \mathbf{i}$, $\lambda \in \mathbb{R}$.

PROOF. By Jensen inequality we have

$$W(x - \bar{\mu}) \leq \int_{\mathbb{R}^d} W(x - y) d\mu(y) \quad \forall x \in \mathbb{R}^d, \quad (4.98)$$

so that a further integration yields

$$\int_{\mathbb{R}^d} W(x - \bar{\mu}) d\mu(x) \leq \mathcal{W}(\mu); \quad (4.99)$$

(4.94) follows directly from (4.99) and the doubling condition (4.92), since $W(x) \leq C_W(1 + W(x - \bar{\mu}) + W(\bar{\mu}))$.

Combining the doubling condition and the convexity of W we also get

$$\begin{aligned} |\nabla W(x)| &= \sup_{|y| \leq 1} \langle \nabla W(x), y \rangle \\ &\leq \sup_{|y| \leq 1} W(x + y) - W(x) \\ &\leq C_W \left(1 + W(x) + \sup_{|y| \leq 1} W(y) \right), \end{aligned} \quad (4.100)$$

which yields (4.95).

If now $\nu \in D(W)$ and $\boldsymbol{\gamma} \in \Gamma(\mu, \nu)$, then the positive part of the map $(x_1, y_1, x_2) \mapsto \langle \nabla W(x_1 - x_2), y_1 - x_1 \rangle$ belongs to $L^1(\boldsymbol{\gamma} \times \mu)$ since convexity yields

$$\langle \nabla W(x_1 - x_2), y_1 - x_1 \rangle \leq W(y_1 - x_2) - W(x_1 - x_2),$$

and the right-hand side of this inequality is integrable:

$$\begin{aligned} \int_{\mathbb{R}^{3d}} W(y_1 - x_2) d\boldsymbol{\gamma} \times \mu &= \int_{\mathbb{R}^{2d}} W(y_1 - x_2) d\nu \times \mu \\ &\leq C(1 + \mathcal{W}(\nu) + \mathcal{W}(\mu) + W(\bar{\nu} - \bar{\mu})), \\ \int_{\mathbb{R}^{3d}} W(x_1 - x_2) d\boldsymbol{\gamma} \times \mu &= \int_{\mathbb{R}^{2d}} W(x_1 - x_2) d\mu \times \mu = \mathcal{W}(\mu). \end{aligned}$$

Therefore we can apply Fubini–Tonelli theorem to obtain

$$\begin{aligned} &\int_{\mathbb{R}^{3d}} \langle \nabla W(x_1 - x_2), y_1 - x_1 \rangle d\boldsymbol{\gamma} \times \mu(x_1, y_1, x_2) \\ &= \int_{\mathbb{R}^{2d}} \left(\int_X \langle \nabla W(x_1 - x_2), y_1 - x_1 \rangle d\mu(x_2) \right) d\boldsymbol{\gamma}(x_1, y_1) \\ &= \int_{\mathbb{R}^{2d}} \left\langle \left(\int_X \nabla W(x_1 - x_2) d\mu(x_2) \right), y_1 - x_1 \right\rangle d\boldsymbol{\gamma}(x_1, y_1) \\ &= \int_{\mathbb{R}^{2d}} \langle \mathbf{w}(x_1), y_1 - x_1 \rangle d\boldsymbol{\gamma}(x_1, y_1), \end{aligned}$$

which yields (4.96). $\square$

As the interaction energy fails to satisfy (4.31a), as we did for the potential energy functional we say that $\boldsymbol{\xi} \in L^2(\mu; \mathbb{R}^d)$ belongs to the Fréchet subdifferential $\partial\mathcal{W}(\mu)$ at $\mu \in D(\mathcal{W})$ if

$$\begin{aligned} &\mathcal{W}(\nu) - \mathcal{W}(\mu) \\ &\geq \inf_{\gamma \in \Gamma_o(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \boldsymbol{\xi}(x), y - x \rangle d\gamma(x) + o(W_2(\mu, \nu)). \end{aligned} \quad (4.101)$$

THEOREM 4.19 (Minimal subdifferential of $\mathcal{W}$). *Assume that $W: \mathbb{R}^d \rightarrow [0, +\infty)$ is convex, Gateaux differentiable, even and satisfies the doubling condition (4.92). Then $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ belongs to $D(|\partial\mathcal{W}|)$ if and only if $\mathbf{w} = (\nabla W) * u \in L^2(\mu; \mathbb{R}^d)$. In this case $\mathbf{w} = \partial^\circ \mathcal{W}(\mu)$.*

PROOF. As we did for the internal energy functional, we start by computing the directional derivative of $\mathcal{W}$ along a direction induced by a transport map $\mathbf{r} = \mathbf{i} + \mathbf{t}$, with $\mathbf{t}$ bounded and with a compact support (by the growth condition on W , this ensures that $\mathcal{W}(\mathbf{r}_\# \mu) < +\infty$). Since the map

$$t \mapsto \frac{W((x - y) + t(\mathbf{t}(x) - \mathbf{t}(y))) - W(x - y)}{t}$$

is nondecreasing w.r.t. t , the monotone convergence theorem and (4.97) give (taking into account that ∇W is an odd function)

$$\begin{aligned} +\infty &> \lim_{t \downarrow 0} \frac{\mathcal{W}((\mathbf{i} + t\mathbf{t})_{\#}\mu) - \mathcal{W}(\mu)}{t} \\ &= \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \nabla W(x - y), (\mathbf{t}(x) - \mathbf{t}(y)) \rangle d\mu \times \mu \\ &= \int_{\mathbb{R}^d} \langle \mathbf{w}, \mathbf{t} \rangle d\mu. \end{aligned}$$

On the other hand, since $|\partial\mathcal{W}|(\mu) < +\infty$, using the inequality $W_2((\mathbf{i} + t\mathbf{t})_{\#}\mu, \mu) \leq \|\mathbf{t}\|_{L^2(\mu; \mathbb{R}^d)}$ we get

$$\int_{\mathbb{R}^d} \langle \mathbf{w}, \mathbf{t} \rangle d\mu \geq -|\partial\mathcal{W}|(\mu) \|\mathbf{t}\|_{L^2(\mu; \mathbb{R}^d)};$$

changing the sign of $\mathbf{t}$ we obtain

$$\left| \int_{\mathbb{R}^d} \langle \mathbf{w}, \mathbf{t} \rangle d\mu \right| \leq |\partial\mathcal{W}|(\mu) \|\mathbf{t}\|_{L^2(\mu; \mathbb{R}^d)},$$

and this proves that $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$ and that $\|\mathbf{w}\|_{L^2} \leq |\partial\mathcal{W}|(\mu)$.

Now we prove that if $\mathbf{w} = (\nabla W) * \mu \in L^2(\mu; \mathbb{R}^d)$, then it belongs to $\partial\mathcal{W}(\mu)$. Let us consider a test measure $\nu \in D(\mathcal{W})$, a plan $\gamma \in \Gamma(\mu, \nu)$, and the directional derivative of $\mathcal{W}$ along the direction induced by γ . Since the map

$$t \mapsto \frac{W((1-t)(x_1 - x_2) + t(y_1 - y_2)) - W(x_1 - x_2)}{t}$$

is nondecreasing w.r.t. t , the monotone convergence theorem, the fact that ∇W is an odd function, and (4.97) give

$$\begin{aligned} \mathcal{W}(\nu) - \mathcal{W}(\mu) &\geq \lim_{t \downarrow 0} \frac{\mathcal{W}(((1-t)\pi^1 + t\pi^2)_{\#}\gamma) - \mathcal{W}(\mu)}{t} \\ &= \frac{1}{2} \int_{\mathbb{R}^{2d} \times \mathbb{R}^{2d}} \langle \nabla W(x_1 - x_2), (y_1 - x_1) - (y_2 - x_2) \rangle d\gamma \times \gamma \\ &= \int_{\mathbb{R}^{2d}} \langle \mathbf{w}(x_1), y_1 - x_1 \rangle d\gamma(x_1, y_1), \end{aligned}$$

and this proves that $\mathbf{w} \in \partial\mathcal{W}(\mu)$. □

4.5.6. The opposite Wasserstein distance. In this section we compute the (metric) slope of the function $\psi(\cdot) := -1/2W_2^2(\cdot, \nu)$, i.e., the limit

$$\frac{1}{2} \limsup_{\sigma \rightarrow \mu} \frac{W_2^2(\sigma, \nu) - W_2^2(\mu, \nu)}{W_2(\sigma, \mu)} = |\partial\psi|(\mu); \quad (4.102)$$

observe that the triangle inequality shows that the “limsup” above is always less than $W_2(\mu, \nu)$; however this inequality is always strict when optimal plans are not induced by transports, as the following theorem shows ([9], Theorem 10.4.12); the right formula for the slope involves the minimal L^2 norm of the *barycentric projection* of the optimal plans and gives that the minimal selection is always induced by a map. We recall that, given $\gamma \in \Gamma(\mu, \nu)$, the barycentric projection $\bar{\gamma}$ is the map in $L^2(\mu)$ characterized by $\pi_{\#}^1(\gamma\gamma) = \bar{\gamma}\mu$, or equivalently by

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} y\varphi(x) d\gamma = \int_{\mathbb{R}^d} \bar{\gamma}(x)\varphi(x) d\mu(x) \quad \forall \varphi \in C_c^\infty(\mathbb{R}^d).$$

THEOREM 4.20 (Minimal subdifferential of $-1/2W_2^2(\cdot, \nu)$). *Let $\psi(\mu) = -1/2W_2^2(\mu, \nu)$. Then*

$$\partial\psi(\mu) = \{\bar{\gamma} - \mathbf{i} : \gamma \in \Gamma_o(\mu, \nu)\} \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d).$$

In particular,

$$|\partial\psi|^2(\mu) = \min \left\{ \int_{\mathbb{R}^d} |\bar{\gamma} - \mathbf{i}|^2 d\mu : \gamma \in \Gamma_o(\mu, \nu) \right\} \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d), \quad (4.103)$$

and $\partial^\circ\psi(\mu) = \bar{\gamma} - \mathbf{i}$ is a strong subdifferential, where γ is the unique minimizing plan above.

Finally, $\mu \mapsto |\partial\psi|(\mu)$ is lower semicontinuous with respect to narrow convergence in $\mathcal{P}(\mathbb{R}^d)$, along sequences bounded in $\mathcal{P}_2(\mathbb{R}^d)$.

4.5.7. The sum of internal, potential and interaction energy. In this section we consider, as in [29], the functional $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ given by the sum of internal, potential and interaction energy:

$$\begin{aligned} \phi(\mu) &:= \int_{\mathbb{R}^d} F(u) dx + \int_{\mathbb{R}^d} V d\mu + \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W d\mu \times \mu \\ &\text{if } \mu = u\mathcal{L}^d, \end{aligned} \quad (4.104)$$

setting $\phi(\mu) = +\infty$ if $\mu \in \mathcal{P}_2(\mathbb{R}^d) \setminus \mathcal{P}_2^a(\mathbb{R}^d)$. Recalling the “doubling condition” stated in (4.77), we make the following assumptions on F , V and W :

- (F) $F : [0, +\infty) \rightarrow \mathbb{R}$ is a doubling, convex differentiable function with superlinear growth satisfying (4.70) (i.e., the bounds on F^-) and (4.75) (yielding the geodesic convexity of the internal energy).

(V) $V : \mathbb{R}^d \rightarrow (-\infty, +\infty]$ is a l.s.c. λ -convex function with proper domain $D(V)$ with nonempty interior $\Omega \subset \mathbb{R}^d$.

(W) $W : \mathbb{R}^d \rightarrow [0, +\infty)$ is a convex, differentiable, even function satisfying the doubling condition (4.92).

The finiteness of ϕ yields

$$\text{supp } \mu \subset \overline{\Omega} = \overline{D(V)}, \quad \mu(\partial\Omega) = 0, \quad (4.105)$$

so that its density u w.r.t. $\mathcal{L}^d$ can be considered as a function of $L^1(\Omega)$.

The same monotonicity argument used in the proof of Lemma 4.14 gives

$$\begin{aligned} +\infty &> \lim_{t \downarrow 0} \frac{\int_{\mathbb{R}^d} V \, d((1-t)\mathbf{i} + t\mathbf{r})_{\#}\mu - \int_{\mathbb{R}^d} V \, d\mu}{t} \\ &= \int_{\mathbb{R}^d} \langle \nabla V, \mathbf{r} - \mathbf{i} \rangle \, d\mu, \end{aligned} \quad (4.106)$$

whenever both $\int_{\mathbb{R}^d} V \, d\mu < +\infty$ and $\int_{\mathbb{R}^d} V \, d\mathbf{r}_{\#}\mu < +\infty$.

Analogously, denoting by $\mathcal{W}$ the interaction energy functional induced by $W/2$, arguing as in the first part of Theorem 4.19 we have

$$\begin{aligned} +\infty &> \lim_{t \downarrow 0} \frac{\mathcal{W}(((1-t)\mathbf{i} + t\mathbf{r})_{\#}\mu) - \mathcal{W}(\mu)}{t} \\ &= \int_{\mathbb{R}^d} \langle (\nabla W) * \mu, \mathbf{r} - \mathbf{i} \rangle \, d\mu, \end{aligned} \quad (4.107)$$

whenever $\mathcal{W}(\mu) + \mathcal{W}(\mathbf{r}_{\#}\mu) < +\infty$. The growth condition on W ensures that $\mu \in D(\mathcal{W})$ implies $\mathbf{r}_{\#}\mu \in D(\mathcal{W})$ if either $\mathbf{r} - \mathbf{i}$ is bounded or $\mathbf{r} = 2\mathbf{i}$ (here we use the doubling condition).

We have the following characterization of the minimal selection in the subdifferential $\partial^\circ \phi(\mu)$.

THEOREM 4.21 (Minimal subdifferential of ϕ). *A measure $\mu = u\mathcal{L}^d \in D(\phi) \subset \mathcal{P}_2(\mathbb{R}^d)$ belongs to $D(|\partial\phi|)$ if and only if $L_F(u) \in W_{\text{loc}}^{1,1}(\Omega)$ and*

$$u\mathbf{w} = \nabla L_F(u) + u\nabla V + u(\nabla W) * u \quad \text{for some } \mathbf{w} \in L^2(\mu; \mathbb{R}^d). \quad (4.108)$$

In this case the vector $\mathbf{w}$ defined μ -a.e. by (4.108) is the minimal selection in $\partial\phi(\mu)$, i.e., $\mathbf{w} = \partial^\circ \phi(\mu)$.

PROOF. We argue exactly as in the proof of Theorem 4.16, computing the Gateaux derivative of ϕ in several directions $\mathbf{r}$, using Lemma 4.14 for the internal energy and (4.106), (4.107) respectively for the potential and interaction energy.

Choosing $\mathbf{r} = \mathbf{i} + \mathbf{t}$, with $\mathbf{t} \in C_c^\infty(\Omega; \mathbb{R}^d)$, we obtain

$$\begin{aligned} & - \int_{\mathbb{R}^d} L_F(u) \nabla \cdot \mathbf{t} \, dx + \int_{\mathbb{R}^d} \langle \nabla V, \mathbf{t} \rangle \, d\mu + \int_{\mathbb{R}^d} \langle (\nabla W) * u, \mathbf{t} \rangle \, d\mu \\ & \geq -|\partial\phi|(\mu) \|\mathbf{t}\|_{L^2(\mu)}. \end{aligned} \quad (4.109)$$

Since V is locally Lipschitz in Ω and $\nabla W * u$ is locally bounded, following the same argument of Theorem 4.16, we obtain from (4.109) first that $L_F(u) \in BV_{\text{loc}}(\mathbb{R}^d)$ and then that $L_F(u) \in W_{\text{loc}}^{1,1}(\mathbb{R}^d)$, with

$$\nabla L_F(u) + u \nabla V + u(\nabla W) * u = \mathbf{w}u \quad \text{for some } \mathbf{w} \in L^2(\mu; \mathbb{R}^d) \quad (4.110)$$

with $\|\mathbf{w}\|_{L^2} \leq |\partial\phi|(\mu)$.

In order to show that the vector $\mathbf{w}$ is in the subdifferential (and then, by the previous estimate, it is the minimal selection) we choose eventually a test measure $\nu \in D(\phi)$ with compact support contained in Ω and the associated optimal transport map $\mathbf{r} = \mathbf{t}_\mu^\nu$; Lemma 4.14, (4.106), (4.107) and Lemma 4.15 yield

$$\begin{aligned} & \phi(\nu) - \phi(\mu) \\ & \geq \frac{d}{dt} \phi((1-t)\mathbf{i} + t\mathbf{r})_\# \mu \Big|_{t=0^+} \\ & = - \int_{\Omega} L_F(u) \nabla \cdot (\mathbf{r} - \mathbf{i}) \, dx + \int_{\Omega} \langle \nabla V, \mathbf{r} - \mathbf{i} \rangle \, d\mu + \int_{\Omega} \langle (\nabla W) * u, \mathbf{r} - \mathbf{i} \rangle \, d\mu \\ & \geq \limsup_{h \rightarrow \infty} \int_{\Omega} \langle \nabla L_F(u), \mathbf{r} - \mathbf{i} \rangle \chi_h \, dx + \int_{\Omega} \langle \nabla V + (\nabla W) * u, \mathbf{r} - \mathbf{i} \rangle \, d\mu \\ & = \limsup_{h \rightarrow \infty} \int_{\Omega} \langle \nabla L_F(u) + u \nabla V + u(\nabla W) * u, \mathbf{r} - \mathbf{i} \rangle \chi_h \, dx \\ & = \int_{\Omega} \langle u \mathbf{w}, \mathbf{r} - \mathbf{i} \rangle \, dx \\ & = \int_{\Omega} \langle \mathbf{w}, \mathbf{r} - \mathbf{i} \rangle \, d\mu. \end{aligned}$$

Finally, we notice that the proof that $\mathbf{w}$ belongs to the subdifferential did not use the finiteness of slope, but only the assumption (previously derived by the finiteness of slope) that $L_F(u) \in W_{\text{loc}}^{1,1}(\Omega)$, (4.108), and $\phi(\mu) < +\infty$; therefore these conditions imply that the subdifferential is not empty, hence the slope is finite and the vector $\mathbf{w}$ is the minimal selection in $\partial\phi(\mu)$. $\square$

An interesting particular case of the above result is provided by the relative entropy functional: let us choose $W \equiv 0$ and

$$F(s) := s \log s, \quad \gamma := \frac{1}{Z} e^{-V} \mathcal{L}^d = e^{-(V(x) + \log Z)} \mathcal{L}^d,$$

with $Z > 0$ chosen so that $\gamma(\mathbb{R}^d) = 1$. Recalling Remark 3.16, the functional ϕ can also be written as

$$\phi(\mu) = \mathcal{H}(\mu|\gamma) - \log Z. \quad (4.111)$$

Since in this case $L_F(u) = u$, a vector $\mathbf{w} \in L^2(\mu; \mathbb{R}^d)$ is the minimal selection $\partial^\circ \phi(\mu)$ if and only if

$$\begin{aligned} & - \int_{\mathbb{R}^d} \nabla \cdot \boldsymbol{\zeta}(x) \, d\mu(x) \\ & = \int_{\mathbb{R}^d} \langle \mathbf{w}(x), \boldsymbol{\zeta}(x) \rangle \, d\mu(x) - \int_{\mathbb{R}^d} \langle \nabla V(x), \boldsymbol{\zeta}(x) \rangle \, d\mu(x), \end{aligned} \quad (4.112)$$

for every test function $\boldsymbol{\zeta} \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$; (4.112) can also be written in terms of $\rho = \frac{d\mu}{d\gamma}$ as

$$- \int_{\mathbb{R}^d} \rho \nabla \cdot (e^{-V(x)} \boldsymbol{\zeta}(x)) \, dx = \int_{\mathbb{R}^d} \langle \rho \mathbf{w}(x), e^{-V(x)} \boldsymbol{\zeta}(x) \rangle \, dx, \quad (4.113)$$

which shows that $\rho \mathbf{w} = \nabla \rho$.

5. Gradient flows of λ -geodesically convex functionals in $\mathcal{P}_2(\mathbb{R}^d)$

In this section we state some structural results, concerning existence, uniqueness, approximation, and qualitative properties of gradient flows in $\mathcal{P}_2(\mathbb{R}^d)$ generated by a proper and l.s.c. functional

$$\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]. \quad (5.1a)$$

We will also assume that

$$\phi \text{ is } \lambda\text{-geodesically convex, according to Definition 3.1.} \quad (5.1b)$$

Since we are mostly concerned with absolutely continuous measures, some technical details will be simpler assuming that

$$D(|\partial \phi|) \subset \mathcal{P}_2^a(\mathbb{R}^d); \quad (5.1c)$$

finally, the (simplified) existence theory we are presenting here will also require that for some $\tau_* > 0$

$$\begin{aligned} & \text{the map } v \mapsto \Phi(\tau, \mu; v) = 1/(2\tau) W_2^2(\mu, v) + \phi(v) \text{ admits at least} \\ & \text{a minimum point } \mu_\tau \text{ for all } \tau \in (0, \tau_*) \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d). \end{aligned} \quad (5.1d)$$

Notice that (5.1c) gives that any minimizer μ_τ in (5.1d) belongs to $\mathcal{P}_2^a(\mathbb{R}^d)$, due to Lemma 4.4.

REMARK 5.1. (5.1d) is slightly more restrictive than lower semicontinuity in $\mathcal{P}_2(\mathbb{R}^d)$; by the standard direct method in Calculus of Variations, it surely holds if ϕ satisfies the following coerciveness-l.s.c. conditions:

$$\inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \phi(\mu) + \frac{1}{2\tau_*} m_2^2(\mu) > -\infty, \quad (5.2a)$$

$$\left\{ \begin{array}{l} \mu_n \rightarrow \mu \text{ narrowly in } \mathcal{P}(\mathbb{R}^d) \\ \sup_n m_2(\mu_n) < +\infty \end{array} \right\} \implies \liminf_{n \rightarrow \infty} \phi(\mu_n) \geq \phi(\mu). \quad (5.2b)$$

Another sufficient condition yielding (5.1d) and satisfied by our main examples is (5.61): it will be introduced in the “existence” Theorem 5.8.

The inclusion (5.1c) is a simplifying assumption, which ensures that the flows stay inside the absolutely continuous measures, thus avoiding more complicated notions of subdifferentials (see Chapter 11 of [9], where this restriction is completely removed).

DEFINITION 5.2 (Gradient flows). We say that $\mu_t \in AC_{\text{loc}}^2((0, +\infty); \mathcal{P}_2(\mathbb{R}^d))$ is a solution of the gradient flow equation

$$\mathbf{v}_t \in -\partial\phi(\mu_t), \quad t > 0, \quad (5.3)$$

if, for $\mathcal{L}^1$ -a.e. $t > 0$, $\mu_t \in \mathcal{P}_2^a(\mathbb{R}^d)$ and its velocity vector field $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ belongs to the subdifferential (4.20) of ϕ at μ_t .

Recalling the characterization of the tangent velocity field to an absolutely continuous curve, the above definition is equivalent to the requirement that there exists a Borel vector field $\mathbf{v}_t$ such that

$$\begin{aligned} \mathbf{v}_t &\in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d) \quad \text{for } \mathcal{L}^1\text{-a.e. } t > 0, \\ \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)} &\in L_{\text{loc}}^2(0, +\infty), \end{aligned} \quad (5.4a)$$

the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty) \quad (5.4b)$$

holds in the sense of distributions according to (2.46), and finally

$$-\mathbf{v}_t \in \partial\phi(\mu_t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t > 0. \quad (5.4c)$$

Before studying the question of existence of solutions to (5.3), which we will postpone to the next sections, we want to discuss some preliminary issues.

5.1. Characterizations of gradient flows, uniqueness and contractivity

THEOREM 5.3 (Gradient flows, EVI, and curves of maximal slope). *Let $\phi: \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ be as in (5.1a) and (5.1b). An absolutely continuous curve $\mu \in AC_{\text{loc}}^2((0, +\infty);$*

$\mathcal{P}_2(\mathbb{R}^d)$) with $\mu_t \in \mathcal{P}_2^a(\mathbb{R}^d)$ for $\mathcal{L}^1$ -a.e. $t \in (0, +\infty)$ is a gradient flow of ϕ according to Definition 5.2 if and only if it satisfies one of the following equivalent characterizations:

(i) There exists a Borel vector field $\tilde{\mathbf{v}}_t$ with $\|\tilde{\mathbf{v}}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$ in $L^2_{\text{loc}}(0, +\infty)$ such that

$$\partial_t \mu_t + \nabla \cdot (\tilde{\mathbf{v}}_t \mu_t) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty), \quad (5.5a)$$

in the sense of distributions, and

$$-\int_{\mathbb{R}^d} \langle \tilde{\mathbf{v}}_t, \mathbf{t}_{\mu_t}^\sigma - \mathbf{i} \rangle d\mu_t \leq \phi(\sigma) - \phi(\mu_t) - \frac{\lambda}{2} W_2^2(\sigma, \mu_t) \quad \forall \sigma \in D(\phi), \quad (5.5b)$$

$\mathcal{L}^1$ -a.e. in $(0, +\infty)$.

(ii) Every Borel vector field $\tilde{\mathbf{v}}_t$ with $\|\tilde{\mathbf{v}}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$ in $L^2_{\text{loc}}(0, +\infty)$ (in particular the velocity vector field $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$) satisfying the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\tilde{\mathbf{v}}_t \mu_t) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty), \quad (5.6a)$$

in the sense of distributions, satisfies the variational inequality

$$-\int_{\mathbb{R}^d} \langle \tilde{\mathbf{v}}_t, \mathbf{t}_{\mu_t}^\sigma - \mathbf{i} \rangle d\mu_t \leq \phi(\sigma) - \phi(\mu_t) - \frac{\lambda}{2} W_2^2(\sigma, \mu_t) \quad \forall \sigma \in D(\phi), \quad (5.6b)$$

for $t \in (0, +\infty) \setminus \mathcal{N}$, $\mathcal{N}$ being an $\mathcal{L}^1$ -negligible set.

(iii) The metric evolution variational inequalities (EVI)

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \sigma) + \frac{\lambda}{2} W_2^2(\mu_t, \sigma) \leq \phi(\sigma) - \phi(\mu_t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t > 0 \quad (5.7)$$

hold for every $\sigma \in D(\phi)$.

(iv) The map $t \mapsto \phi(\mu_t)$ is locally absolutely continuous in $(0, +\infty)$ and

$$-\frac{d}{dt} \phi(\mu_t) \geq \frac{1}{2} \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)}^2 + \frac{1}{2} |\partial \phi|^2(\mu_t) \quad \mathcal{L}^1\text{-a.e. in } (0, +\infty). \quad (5.8)$$

(v) The map $t \mapsto \phi(\mu_t)$ is locally absolutely continuous in $(0, +\infty)$ and

$$-\frac{d}{dt} \phi(\mu_t) = \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)}^2 = |\partial \phi|^2(\mu_t) \quad \mathcal{L}^1\text{-a.e. in } (0, +\infty). \quad (5.9)$$

In particular, (5.3) and (v) yield

$$-\mathbf{v}_t = \partial^\circ \phi(\mu_t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t > 0. \quad (5.10)$$

PROOF. (i) If μ_t is a gradient flow according to Definition 5.2, recalling the property of the subdifferential (4.37), it is immediate that μ_t and its velocity vector field $\mathbf{v}_t$ satisfy (5.5a) and (5.5b).

Conversely, suppose that $\tilde{\mathbf{v}}_t$ satisfies (5.5a) and (5.5b) and let us denote by $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ the tangent velocity vector of μ_t . Since, by (2.55), for $\mathcal{L}^1$ -a.e. $t > 0$ $\mathbf{v}_t$ is the orthogonal projection of $\tilde{\mathbf{v}}_t$ on $\text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$, the difference $\tilde{\mathbf{v}}_t - \mathbf{v}_t$ is orthogonal to the tangent space, and therefore by Theorem 2.22 we have

$$\int_{\mathbb{R}^d} \langle \tilde{\mathbf{v}}_t - \mathbf{v}_t, \mathbf{t}_{\mu_t}^\sigma - \mathbf{i} \rangle d\mu_t = 0 \quad \forall \sigma \in \mathcal{P}_2(\mathbb{R}^d), \text{ for } \mathcal{L}^1\text{-a.e. } t > 0. \quad (5.11)$$

As a consequence, $\mathbf{v}_t$ fulfills (5.5b) for $\mathcal{L}^1$ -a.e. t , and this property characterizes the elements of the subdifferential.

(ii) Follows by the same argument, thanks to (5.11).

(iii) Assume that (5.7) holds for all $\sigma \in D(\phi)$. For any $\sigma \in D(\phi)$ fixed, the differentiability of W_2^2 stated in Lemma 2.21 gives

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \sigma) = \int_{\mathbb{R}^d} \langle \mathbf{v}_t, \mathbf{i} - \mathbf{t}_{\mu_t}^\sigma \rangle d\mu_t \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, +\infty).$$

Therefore we can find, for any countable set $\mathcal{D} \subset D(\phi)$, an $\mathcal{L}^1$ -negligible set of times $\mathcal{N}$ such that

$$-\int_{\mathbb{R}^d} \langle \mathbf{v}_t, \mathbf{t}_{\mu_t}^\sigma - \mathbf{i} \rangle d\mu_t \leq \phi(\sigma) - \phi(\mu_t) - \frac{\lambda}{2} W_2^2(\sigma, \mu_t) \quad (5.12)$$

holds for all $t \in (0, +\infty) \setminus \mathcal{N}$ and all $\phi \in \mathcal{D}$. Choosing $\mathcal{D}$ to be dense relative to the distance $W_2(\mu, \nu) + |\phi(\mu) - \phi(\nu)|$ in $D(\phi)$, we obtain that (5.5b) holds for all $t \in (0, +\infty) \setminus \mathcal{N}$. The converse implication is analogous.

(iv) If μ_t is a gradient flow in the sense of (5.3), taking into account that $|\mu_t'| = \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$ and that $|\partial\phi(\mu_t)| \leq \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$ (by (4.52)) we obtain

$$|\partial\phi(\mu_t)| |\mu_t'| \in L_{\text{loc}}^1(0, +\infty).$$

Thanks to the λ -convexity and the lower semicontinuity of ϕ , this implies (see (4.56) or Corollary 2.4.10 in [9]) that $t \mapsto \phi(\mu_t)$ is locally absolutely continuous in $(0, +\infty)$. Then, the chain rule (4.55) easily yields

$$-\frac{d}{dt} \phi(\mu_t) = \int_{\mathbb{R}^d} |\mathbf{v}_t|^2 d\mu_t \geq |\partial\phi|^2(\mu_t) \quad (5.13)$$

for $\mathcal{L}^1$ -a.e. $t > 0$, and therefore (5.8).

Conversely, if $t \mapsto \phi(\mu_t)$ is locally absolutely continuous and μ_t satisfies (5.8), we know that $\partial\phi(\mu_t) \neq \emptyset$ for $\mathcal{L}^1$ -a.e. $t > 0$; thus the chain rule (4.55) shows that

$$\frac{d}{dt} \phi(t) = \int_{\mathbb{R}^d} \langle \xi, \mathbf{v}_t \rangle d\mu_t \quad \forall \xi \in \partial\phi(\mu_t), \text{ for } \mathcal{L}^1\text{-a.e. } t > 0. \quad (5.14)$$

Choosing in particular $\xi_t = \partial^\circ \phi(\mu_t)$, for $\mathcal{L}^1$ -a.e. $t > 0$ we get

$$\int_{\mathbb{R}^d} \left(\frac{1}{2} |\mathbf{v}_t|^2 + \frac{1}{2} |\xi_t|^2 + \langle \xi_t, \mathbf{v}_t \rangle \right) d\mu_t \leq 0. \quad (5.15)$$

It follows that

$$\xi_t(x) = -\mathbf{v}_t(x) \quad \text{for } \mu_t\text{-a.e. } x \in \mathbb{R}^d,$$

i.e., $\mathbf{v}_t = -\partial^\circ \phi(\mu_t)$.

(v) is equivalent to (iv) by the previous argument. $\square$

REMARK 5.4. The “purely metric” formulations (5.7) or (5.8) do not require that μ_t is an absolutely continuous measure at $\mathcal{L}^1$ -a.e. $t \in (0, +\infty)$ and do not depend on an explicit expression of the subdifferential of ϕ , as only the metric slope is involved; therefore they can be used to define the gradient flow of ϕ under more general assumptions: again, we refer to [9] for a complete development of this approach. Different points of view have been considered in [29,76].

THEOREM 5.5 (Uniqueness and contractivity of gradient flows). *If $\mu_t^i : (0, +\infty) \rightarrow \mathcal{P}_2(\mathbb{R}^d)$, $i = 1, 2$, are gradient flows satisfying $\mu_t^i \rightarrow \mu^i \in \mathcal{P}_2(\mathbb{R}^d)$ as $t \downarrow 0$ in $\mathcal{P}_2(\mathbb{R}^d)$, then*

$$W_2(\mu_t^1, \mu_t^2) \leq e^{-\lambda t} W_2(\mu^1, \mu^2) \quad \forall t > 0. \quad (5.16)$$

In particular, for any $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ there is at most one gradient flow μ_t satisfying the initial Cauchy condition $\mu_t \rightarrow \mu_0$ as $t \downarrow 0$.

PROOF. If μ_t^1, μ_t^2 are two gradient flows satisfying the initial Cauchy condition $\mu_t^i \rightarrow \mu^i$ as $t \downarrow 0$, $i = 1, 2$, by the EVI formulation (5.7) we can apply the next Lemma 5.6 with the choices $d(s, t) := W_2^2(\mu_s^1, \mu_t^2)$, $\delta(t) := d(t, t)$, thus obtaining $\delta' \leq -2\lambda\delta$. Since $\delta(0_+) = W_2^2(\mu^1, \mu^2)$ we obtain (5.16). $\square$

LEMMA 5.6. *Let $d(s, t) : (a, b)^2 \rightarrow \mathbb{R}$ be a map satisfying*

$$|d(s, t) - d(s', t)| \leq |v(s) - v(s')|, \quad |d(s, t) - d(s, t')| \leq |v(t) - v(t')|$$

for any $s, t, s', t' \in (a, b)$, for some locally absolutely continuous map $v : (a, b) \rightarrow \mathbb{R}$ and let $\delta(t) := d(t, t)$. Then δ is locally absolutely continuous in (a, b) and

$$\begin{aligned} \frac{d}{dt} \delta(t) &\leq \limsup_{h \downarrow 0} \frac{d(t, t) - d(t-h, t)}{h} + \limsup_{h \downarrow 0} \frac{d(t, t+h) - d(t, t)}{h} \\ &\mathcal{L}^1\text{-a.e. in } (a, b). \end{aligned}$$

PROOF. Since $|\delta(s) - \delta(t)| \leq 2|v(s) - v(t)|$ the function δ is locally absolutely continuous. We fix a nonnegative function $\zeta \in C_c^\infty(a, b)$ and $h > 0$ such that $\pm h + \text{supp } \zeta \subset (a, b)$. We have then

$$\begin{aligned} & - \int_a^b \delta(t) \frac{\zeta(t+h) - \zeta(t)}{h} dt \\ &= \int_a^b \zeta(t) \frac{d(t, t) - d(t-h, t-h)}{h} dt \\ &= \int_a^b \zeta(t) \frac{d(t, t) - d(t-h, t)}{h} dt + \int_a^b \zeta(t+h) \frac{d(t, t+h) - d(t, t)}{h} dt, \end{aligned}$$

where the last equality follows by adding and subtracting $d(t-h, t)$ and then making a change of variables in the last integral. Since

$$\begin{aligned} h^{-1} |d(t, t) - d(t-h, t)| &\leq h^{-1} |v(t) - v(t-h)| \rightarrow |v'(t)| \\ &\text{in } L_{\text{loc}}^1(a, b) \text{ and pointwise } \mathcal{L}^1\text{-a.e. in } (a, b) \text{ as } h \downarrow 0 \end{aligned}$$

and an analogous inequality holds for the other difference quotient, we can apply (an extended version of) Fatou's lemma and pass to the upper limit in the integrals as $h \downarrow 0$ (recall that Fatou's lemma with the limsup holds even for sequences bounded above by a sequence converging both pointwise a.e. and strongly in L^1); denoting by α and β the two upper derivatives in the statement of the lemma we get $-\int \delta \zeta' dt \leq \int (\alpha + \beta) \zeta dt$, whence the inequality between distributions follows. $\square$

5.2. Main properties of gradient flows

In this section we collect the main properties of the gradient flow generated by a functional $\phi: \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ satisfying the assumptions (5.1a)–(5.1c). We limit this exposition to functionals ϕ whose modulus of (geodesic) convexity is quadratic (λ -convexity according to Definition 3.1); more general assumptions could also be considered as in [29].

THEOREM 5.7 (Main properties of gradient flows). *Let us suppose that $\phi: \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ satisfies (5.1a)–(5.1c) and let us suppose that its gradient flow μ_t exists for every initial value $\mu_0 \in D$, D being a dense subset of $D(\phi)$.*

λ -contractive semigroup. For every $\mu_0 \in \overline{D(\phi)}$ there exists a unique solution $\mu := S[\mu_0]$ of the Cauchy problem associated to (5.3) with $\lim_{t \downarrow 0} \mu_t = \mu_0$. The map $\mu_0 \mapsto S_t[\mu_0]$ is a λ -contracting semigroup on $\overline{D(\phi)}$, i.e.,

$$W_2(S[\mu_0](t), S[v_0](t)) \leq e^{-\lambda t} W_2(\mu_0, v_0) \quad \forall \mu_0, v_0 \in \overline{D(\phi)}. \quad (5.17)$$

Regularizing effect. S_t maps $\overline{D(\phi)}$ into $D(\partial\phi) \subset D(\phi)$ for every $t > 0$,

$$\text{the map } t \mapsto e^{\lambda t} |\partial\phi|(\mu_t) \text{ is nonincreasing,} \quad (5.18)$$

and each solution $\mu_t = S_t[\mu_0]$ satisfies the following regularization estimates:

$$\begin{cases} \phi(\mu_t) \leq \frac{1}{2t} W_2^2(\mu_0, \sigma) + \phi(v) & \text{if } \lambda = 0, \\ \phi(\mu_t) \leq \frac{\lambda}{2(e^{\lambda t} - 1)} W_2^2(\mu_0, \sigma) + \phi(v) & \text{if } \lambda \neq 0, \end{cases} \quad (5.19)$$

$$\begin{aligned} & e^{-2\lambda t} |\partial\phi|^2(\mu_t) \\ & \leq |\partial\phi|^2(v) - \frac{\lambda}{2t} W_2^2(\mu_t, \sigma) + \frac{1}{t^2} W_2^2(\mu_0, v) - \frac{\lambda}{t^2} \int_0^t W_2^2(\mu_s, v) \, ds \end{aligned} \quad (5.20)$$

for every $\sigma \in D(\partial\phi)$.

Energy identity. If $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ is the tangent velocity field of a gradient flow $\mu_t = S_t[\mu_0]$, then the energy identity holds:

$$\int_a^b \int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 \, d\mu_t(x) \, dt + \phi(\mu_b) = \phi(\mu_a) \quad \forall 0 \leq a < b < +\infty. \quad (5.21)$$

Asymptotic behavior. If $\lambda > 0$, then ϕ admits a unique minimum point $\bar{\mu}$ and for $t \geq t_0$ we have

$$\frac{\lambda}{2} W_2^2(\mu_t, \bar{\mu}) \leq \phi(\mu_t) - \phi(\bar{\mu}) \leq \frac{1}{2\lambda} |\partial\phi|^2(\mu_t) \quad \forall t \geq 0, \quad (5.22a)$$

$$W_2(\mu_t, \bar{\mu}) \leq W_2(\mu_{t_0}, \bar{\mu}) e^{-\lambda(t-t_0)}, \quad (5.22b)$$

$$\phi(\mu_t) - \phi(\bar{\mu}) \leq (\phi(\mu_{t_0}) - \phi(\bar{\mu})) e^{-2\lambda(t-t_0)}, \quad (5.22c)$$

$$|\partial\phi|(\mu_t) \leq |\partial\phi|(\mu_{t_0}) e^{-\lambda(t-t_0)}. \quad (5.22d)$$

If $\lambda = 0$ and $\bar{\mu}$ is any minimum point of ϕ then we have

$$|\partial\phi|(\mu_t) \leq \frac{W_2(\mu_0, \bar{\mu})}{t}, \quad \phi(\mu_t) - \phi(\bar{\mu}) \leq \frac{W_2^2(\mu_0, \bar{\mu})}{2t}, \quad (5.23)$$

the map $t \mapsto W_2(\mu_t, \bar{\mu})$ is not increasing.

Right and left limits, precise pointwise formulation of the equation. For every $t > 0$ the right limit

$$\mathbf{v}_{t+} := \lim_{h \downarrow 0} \frac{\mathbf{t}_{\mu_t}^{\mu_{t+h}} - \mathbf{i}}{h} \quad \text{exists in } L^2(\mu_t; \mathbb{R}^d) \quad (5.24)$$

and satisfies

$$-\mathbf{v}_{t+} = \partial^\circ \phi(\mu_t) \quad \forall t > 0, \quad (5.25)$$

$$\frac{d}{dt_+} \phi(\mu_t) = - \int_{\mathbb{R}^d} |\mathbf{v}_{t+}|^2 \, d\mu_t = -|\partial\phi|^2(\mu_t) \quad \forall t > 0. \quad (5.26)$$

(5.24), (5.25) and (5.26) hold at $t = 0$ iff $\mu_0 \in D(\partial\phi) = D(|\partial\phi|)$. Moreover, there exists an at most countable set $\mathcal{C} \subset (0, +\infty)$ such that the analogous identities for the left limits hold for every $t \in (0, +\infty) \setminus \mathcal{C}$,

$$\begin{cases} \mathbf{v}_{t-} = \lim_{h \downarrow 0} \frac{\mathbf{t}_{\mu_t}^{\mu_{t-h}} - \mathbf{i}}{h} = -\partial^\circ \phi(\mu_t), \\ \frac{d}{dt} \phi(\mu_t) = -|\partial\phi|^2(\mu_t), \end{cases} \quad \forall t \in (0, +\infty) \setminus \mathcal{C}. \quad (5.27)$$

PROOF. *Regularizing effect.* We first observe that for every $h > 0$ the map $t \mapsto \mu_{t+h}$ is still a gradient flow, and therefore estimate (5.16) yields

$$W_2(\mu_{t+h}, \mu_t) \leq e^{-\lambda(t-t_0)} W_2(\mu_{t_0+h}, \mu_{t_0}) \quad \forall 0 \leq t_0 < t < +\infty. \quad (5.28)$$

Setting

$$\delta(t) := \limsup_{h \downarrow 0} \frac{W_2(\mu_{t+h}, \mu_t)}{h}, \quad t \geq 0, \quad (5.29)$$

(5.28) yields

$$\text{the map } t \mapsto e^{\lambda t} \delta(t) \text{ is nonincreasing.} \quad (5.30)$$

We denote by $\mathcal{N}$ the subset of $(0, +\infty)$ whose points t_0 satisfies $\mu_{t_0} \in D(\partial\phi) \subset \mathcal{P}_2^a(\mathbb{R}^d)$, the metric derivative of μ_t coincides with $\|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$ and $-\mathbf{v}_{t_0} = \partial^\circ \phi(\mu_{t_0})$: by the definition of gradient flow, Theorem 2.15, and point (v) of Theorem 5.3, $\mathcal{L}^1((0, +\infty) \setminus \mathcal{N}) = 0$ and

$$\delta(t) = \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)} = |\partial\phi|(\mu_t) < +\infty \quad \forall t \in (0, +\infty) \setminus \mathcal{N}; \quad (5.31)$$

in particular, (5.30) yields $\delta(t) < +\infty$ for every $t > 0$.

We want to show now that

$$\delta(t) = |\partial\phi|(\mu_t) \quad \forall t \geq 0. \quad (5.32)$$

Integrating the EVI (5.7) in the interval $(t, t+h)$ and dividing by h we get, for every $\sigma \in D(\phi)$,

$$\begin{aligned} & \frac{1}{h} \int_0^h (\phi(\mu_{t+s}) + \frac{\lambda}{2} W_2^2(\mu_{t+s}, \sigma)) ds - \phi(\sigma) \\ & \leq \frac{1}{2h} W_2^2(\mu_t, \sigma) - \frac{1}{2h} W_2^2(\mu_{t+h}, \sigma) \\ & \leq \frac{W_2(\mu_{t+h}, \mu_t)}{2h} (W_2(\mu_t, \sigma) + W_2(\mu_{t+h}, \sigma)). \end{aligned} \quad (5.33)$$

Passing to the limit as $h \downarrow 0$ and recalling that the map $t \mapsto \phi(\mu_t)$ is (absolutely) continuous, we obtain

$$\phi(\mu_t) - \phi(\sigma) + \frac{\lambda}{2} W_2^2(\mu_t, \sigma) \leq \delta(t) W_2(\mu_t, \sigma), \quad (5.34)$$

which yields

$$|\partial\phi|(\mu_t) \leq \delta(t) \quad \forall t \geq 0. \quad (5.35)$$

Choosing $\sigma := \mu_t$ in (5.33), and rescaling the integrand, we can use (4.40) to obtain

$$\begin{aligned} & \frac{1}{2h^2} W_2^2(\mu_{t+h}, \mu_t) \\ & \leq \frac{1}{h} \int_0^1 \left(\phi(\mu_t) - \phi(\mu_{t+hs}) - \frac{\lambda}{2} W_2^2(\mu_{t+hs}, \mu_t) \right) ds \\ & \leq |\partial\phi|(\mu_t) \int_0^1 \frac{W_2(\mu_{t+hs}, \mu_t)}{hs} s ds - \lambda \int_0^1 \frac{W_2^2(\mu_{t+hs}, \mu_t)}{h} ds. \end{aligned}$$

Passing to the limit as $h \downarrow 0$ we obtain

$$\frac{1}{2} \delta^2(t) \leq |\partial\phi|(\mu_t) \int_0^1 \delta(t) s ds = \frac{1}{2} |\partial\phi|(\mu_t) \delta(t), \quad (5.36)$$

which yields (5.32) and in particular (5.18).

The estimates (5.19) follow easily by integrating in the interval $(0, t)$ the following form of (5.7)

$$\frac{d}{ds} \frac{e^{\lambda s}}{2} W_2^2(\mu_s, \sigma) + e^{\lambda s} \phi(\mu_s) \leq e^{\lambda s} \phi(\sigma) \quad (5.37)$$

and recalling that $t \mapsto \phi(\mu_t)$ is nonincreasing; when $\lambda \neq 0$ we get

$$\begin{aligned} \frac{e^{\lambda t} - 1}{\lambda} \phi(\mu_t) & \leq - \int_0^t \frac{d}{ds} \frac{e^{\lambda s}}{2} W_2^2(\mu_s, \sigma) ds + \int_0^t e^{\lambda s} \phi(\sigma) ds \\ & \leq \frac{1}{2} W_2^2(\mu_0, \sigma) + \frac{e^{\lambda t} - 1}{\lambda} \phi(\sigma). \end{aligned}$$

In order to show (5.20) we apply (5.18), the fact that $-\frac{d}{dt} \phi(\mu_t) = |\partial\phi|^2(\mu_t)$ and finally the EVI to obtain

$$\begin{aligned} & \frac{e^{-2\lambda^- t} t^2}{2} |\partial\phi|^2(\mu_t) \\ & \leq \int_0^t s e^{-2\lambda^- s} |\partial\phi|^2(\mu_s) ds \leq - \int_0^t s (\phi(\mu_s))' ds \end{aligned}$$

$$\begin{aligned}
&= \int_0^t \phi(\mu_s) ds - t\phi(\mu_t) \\
&\leq t(\phi(\sigma) - \phi(\mu_t)) + \frac{1}{2}W_2^2(\mu_0, \sigma) - \frac{1}{2}W_2^2(\mu_t, \sigma) - \frac{\lambda}{2} \int_0^t W_2^2(\mu_s, \sigma) ds.
\end{aligned}$$

If $\sigma \in D(\partial\phi)$, using (4.40) we can bound the right-hand side by

$$\begin{aligned}
&t|\partial\phi|(\sigma)W_2(\mu_t, \sigma) \\
&\quad - \frac{1}{2}(t\lambda + 1)W_2^2(\mu_t, \sigma) + \frac{1}{2}W_2^2(\mu_0, \sigma) - \frac{\lambda}{2} \int_0^t W_2^2(\mu_s, \sigma) ds \\
&\leq \frac{t^2}{2}|\partial\phi|^2(\sigma) - \frac{t\lambda}{2}W_2^2(\mu_t, \sigma) + \frac{1}{2}W_2^2(\mu_0, \sigma) - \frac{\lambda}{2} \int_0^t W_2^2(\mu_s, \sigma) ds,
\end{aligned}$$

which yields (5.20).

λ -contractive semigroup. Thanks to the λ -contraction estimate of Theorem 5.5, it is now easy to extend the semigroup S defined on D to its closure, which coincides with $\overline{D(\phi)}$. Observe that each trajectory μ_t of the extended semigroup still satisfies the EVI formulation (5.7); moreover, the previous regularization estimates show that $t \mapsto \mu_t$ is locally Lipschitz and $\mu_t \in D(|\partial\phi|)$ for every $t > 0$, in particular $\mu_t \in \mathcal{P}_2^a(\mathbb{R}^d)$ for every $t > 0$. Theorem 5.3 then shows that μ_t is a gradient flow for ϕ .

Energy identity. It is an immediate consequence of (5.9).

Asymptotic behavior. When $\lambda > 0$ (5.28) shows that for every gradient flow μ_t the sequence $k \mapsto \mu_k$ satisfies the Cauchy condition in $\mathcal{P}_2(\mathbb{R}^d)$, since

$$W_2(\mu_{k+1}, \mu_k) \leq e^{-\lambda} W_2(\mu_k, \mu_{k-1}). \quad (5.38)$$

Therefore it is convergent to some limit $\bar{\mu}$; (5.19) and the lower semicontinuity of ϕ show that $\bar{\mu}$ is a minimum point for ϕ ; in particular, the constant curve $t \mapsto \bar{\mu}$ is a gradient flow. (5.22b) is a particular case of the λ -contraction property (5.17) and in particular it shows that the minimum point $\bar{\mu}$ is unique, when $\lambda > 0$.

The inequality (5.22d) is simply (5.30), while (5.22a) is a general property of λ -geodesically convex functions (even in metric spaces, see Theorem 2.4.14 of [9]): in fact, if $\mu \in D(\partial\phi)$, property (4.40) of the slope and Young inequality yield

$$\phi(\mu) - \phi(\bar{\mu}) \leq |\partial\phi|(\mu)W_2(\mu, \bar{\mu}) - \frac{\lambda}{2}W_2^2(\mu, \bar{\mu}) \leq \frac{1}{2\lambda}|\partial\phi|^2(\mu). \quad (5.39)$$

For the opposite inequality, being $0 \in \partial\phi(\bar{\mu})$, from (4.37) we easily get

$$\phi(\mu) - \phi(\bar{\mu}) \geq \frac{\lambda}{2}W_2^2(\mu, \bar{\mu}). \quad (5.40)$$

The estimate (5.22c) now follows by observing that (5.39) yields

$$\frac{d}{dt}(\phi(\mu_t) - \phi(\bar{\mu})) = -|\partial\phi|^2(\mu_t) \leq -2\lambda(\phi(\mu_t) - \phi(\bar{\mu})). \quad (5.41)$$

Right and left limits, precise pointwise formulation of the equation. Here, for the sake of simplicity, we are assuming that $\lambda \geq 0$.

We already know that $\partial\phi(\mu_t)$ is not empty for $t > 0$: we set $\xi_t = \partial^\circ\phi(\mu_t)$; since the slope $|\partial\phi|$ is lower semicontinuous (see (4.43)) and the map $t \mapsto |\partial\phi|(\mu_t)$ is nonincreasing, we obtain

$$|\partial\phi|(\mu_t) = \lim_{h \downarrow 0} |\partial\phi|(\mu_{t+h}). \quad (5.42)$$

Moreover, the map $t \mapsto \phi(\mu_t)$ is absolutely continuous, nonincreasing, and its time derivative coincides $\mathcal{L}^1$ -a.e. with the nondecreasing map $-|\partial\phi|^2(\mu_t)$; it follows that $t \mapsto \phi(\mu_t)$ is continuous and convex, so that

$$\exists \frac{d}{dt+} \phi(\mu_t) = \lim_{h \downarrow 0} \frac{\phi(\mu_{t+h}) - \phi(\mu_t)}{h} = -|\partial\phi|^2(\mu_t) = -\delta^2(t) \quad \forall t > 0. \quad (5.43)$$

Let now fix $t > 0$ and an infinitesimal sequence h_n such that

$$\frac{\mathbf{t}_{\mu_t}^{\mu_{t+h_n}} - \mathbf{i}}{h_n} \rightharpoonup \tilde{\mathbf{v}}_t \quad \text{weakly in } L^2(\mu_t; \mathbb{R}^d). \quad (5.44)$$

By the definition of subdifferential, it is immediate to check that

$$-|\partial\phi|^2(\mu_t) = -\|\xi_t\|_{L^2(\mu_t; \mathbb{R}^d)}^2 = \frac{d}{dt+} \phi(\mu_t) \geq \int_{\mathbb{R}^d} \langle \xi_t, \tilde{\mathbf{v}}_t \rangle d\mu_t. \quad (5.45)$$

On the other hand,

$$\|\tilde{\mathbf{v}}_t\|_{L^2(\mu_t; \mathbb{R}^d)} \leq \delta(t) = \|\xi_t\|_{L^2(\mu_t; \mathbb{R}^d)}. \quad (5.46)$$

It follows that $\tilde{\mathbf{v}}_t = -\xi_t$; since the limit is uniquely determined independently of the subsequence h_n , we obtain that

$$\lim_{h \downarrow 0} \frac{\mathbf{t}_{\mu_t}^{\mu_{t+h}} - \mathbf{i}}{h} = -\xi_t \quad \text{weakly in } L^2(\mu_t; \mathbb{R}^d). \quad (5.47)$$

On the other hand,

$$\limsup_{h \downarrow 0} \left\| \frac{\mathbf{t}_{\mu_t}^{\mu_{t+h}} - \mathbf{i}}{h} \right\|_{L^2(\mu_t; \mathbb{R}^d)} = \limsup_{h \downarrow 0} \frac{W_2(\mu_t, \mu_{t+h})}{h} = \delta(t) = \|\xi_t\|_{L^2(\mu_t; \mathbb{R}^d)}$$

and therefore the limit in (5.47) is also *strong* in $L^2(\mu_t; \mathbb{R}^d)$.

The same argument can be applied for the left limit at each continuity point of the map $t \mapsto |\partial\phi|(\mu_t)$ (whose complement $\mathcal{C}$ in $(0, +\infty)$ is at most countable), i.e., for every t such that

$$\lim_{h \downarrow 0} |\partial\phi|(\mu_{t-h}) = |\partial\phi|(\mu_t), \quad (5.48)$$

observing that in this case

$$\exists \frac{d}{dt} \phi(\mu_t) = -|\partial \phi|^2(\mu_t) \quad (5.49)$$

and (by the $\mathcal{L}^1$ -a.e. equality of $|\mu_t|'$ and $|\partial \phi(\mu_t)|$ and the monotonicity of $|\partial \phi(\mu_t)|$)

$$\frac{W_2(\mu_{t-h}, \mu_t)}{h} \leq \frac{1}{h} \int_{t-h}^t |\mu_s'| ds = \frac{1}{h} \int_{t-h}^t |\partial \phi|(\mu_s) ds \leq |\partial \phi|(\mu_{t-h}), \quad (5.50)$$

and therefore for any $t \in (0, +\infty) \setminus \mathcal{C}$ we have

$$\limsup_{h \downarrow 0} \left\| \frac{\mathbf{t}_{\mu_t}^{\mu_{t-h}} - \mathbf{i}}{h} \right\|_{L^2(\mu_t; \mathbb{R}^d)} = \limsup_{h \downarrow 0} \frac{W_2(\mu_{t-h}, \mu_t)}{h} \leq |\partial \phi|(\mu_t). \quad (5.51)$$

□

5.3. Existence of gradient flows by convergence of the “minimizing movement” scheme

The existence of solutions to the Cauchy problem for (5.3) will be obtained as limit of a variational approximation scheme (the “minimizing movement” scheme, in De Giorgi’s terminology [36]), which we will briefly recall.

The variational approximation scheme. Let us introduce a uniform partition $\mathcal{P}_\tau$ of $(0, +\infty)$ by intervals I_τ^n of size $\tau > 0$

$$\begin{aligned} \mathcal{P}_\tau &:= \{0 < t_\tau^1 = \tau < t_\tau^2 = 2\tau < \dots < t_\tau^n = n\tau < \dots\}, \\ I_\tau^n &:= ((n-1)\tau, n\tau], \end{aligned}$$

and a given family of “discrete” values M_τ^0 approximating the initial value $\mu_0 \in \overline{D(\phi)}$ so that

$$M_\tau^0 \rightarrow \mu_0 \quad \text{in } \mathcal{P}_2(\mathbb{R}^d), \quad \phi(M_\tau^0) \rightarrow \phi(\mu_0) \quad \text{as } \tau \downarrow 0. \quad (5.52)$$

If (5.1c) and (5.1d) are satisfied, for every $\tau \in (0, \tau_*)$ we can find sequences $(M_\tau^n)_{n \in \mathbb{N}} \subset \mathcal{P}_2^a(\mathbb{R}^d)$ recursively defined by solving the variational problem

$$M_\tau^n \text{ minimizes } \mu \mapsto \Phi(\tau, M_\tau^{n-1}; \mu) = \frac{1}{2\tau} W_2^2(\mu, M_\tau^{n-1}) + \phi(\mu). \quad (5.53)$$

We call “discrete solution” the piecewise constant interpolant

$$\bar{M}_\tau(t) := M_\tau^n \quad \text{if } t \in ((n-1)\tau, n\tau], \quad (5.54)$$

and we say that a curve μ_t is a *minimizing movement* of Φ starting from μ_0 , writing $\mu_t \in MM(\Phi; \mu_0)$, if there exists a family of discrete solutions $\bar{M}_\tau$ such that

$$\bar{M}_\tau(t) \rightarrow \mu_t \quad \text{in } \mathcal{P}_2(\mathbb{R}^d) \quad \text{for every } t > 0, \text{ as } \tau \downarrow 0. \quad (5.55)$$

In order to clarify why this variational scheme provides an approximation of the gradient flow equation (5.3), we introduce the optimal transport maps $\mathbf{t}_\tau^n = \mathbf{t}_{M_\tau^n}^{M_\tau^{n-1}}$ pushing M_τ^n to M_τ^{n-1} , and we define the discrete velocity vector $\mathbf{V}_\tau^n$ as $(\mathbf{i} - \mathbf{t}_\tau^n)/\tau$. By Lemma 4.4,

$$-\mathbf{V}_\tau^n = \frac{\mathbf{t}_\tau^n - \mathbf{i}}{\tau} \in \partial\phi(M_\tau^n), \quad (5.56)$$

which can be considered as an Euler implicit discretization of (5.3). By introducing the piecewise constant interpolant

$$\bar{\mathbf{V}}_\tau(t) := \mathbf{V}_\tau^n \quad \text{if } t \in ((n-1)\tau, n\tau], \quad (5.57)$$

the identity (5.56) reads

$$-\bar{\mathbf{V}}_\tau(t) \in \partial\phi(\bar{M}_\tau(t)) \quad \text{for } t > 0. \quad (5.58)$$

By general compactness arguments, it is not difficult to show that, up to subsequences, $\bar{\mathbf{V}}_\tau \bar{M}_\tau \rightharpoonup \mathbf{v}\mu$ in the distribution sense in $\mathbb{R}^d \times (0, +\infty)$, for some vector field $\mathbf{v}(x, t) = \mathbf{v}_t(x)$ satisfying

$$\begin{aligned} \partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) &= 0 \\ \text{in } \mathbb{R}^d \times (0, +\infty), \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)} &\in L^2_{\text{loc}}(0, +\infty). \end{aligned} \quad (5.59)$$

The main difficulty is to show that the nonlinear equation (5.58) is preserved in the limit.

Here we present two proofs of this fact based on two qualitatively different assumptions: the first one is a *coercivity assumption*: for every $C > 0$ the sublevels

$$\{\mu \in \mathcal{P}_2(\mathbb{R}^d): \phi(\mu) \leq C, m_2(\mu) \leq C\} \quad \text{are compact in } \mathcal{P}_2(\mathbb{R}^d). \quad (5.60)$$

The second one is a *strong convexity assumption*: for every $\mu \in D(|\partial\phi|)$ and $\sigma_0, \sigma_1 \in D(\phi)$

$$\text{the map} \quad \begin{cases} s \mapsto \phi(\sigma_s) - \frac{\lambda}{2} W_2^2(\sigma_0, \sigma_1) s^2 \\ \sigma_s := ((1-s)\mathbf{t}_\mu^{\sigma_0} + s\mathbf{t}_\mu^{\sigma_1})_\# \mu \end{cases} \quad \text{is convex in } [0, 1]. \quad (5.61)$$

The first assumption is typically satisfied when the domain of ϕ consists of measures supported in a bounded domain (as in this case convergence in $\mathcal{P}_2(\mathbb{R}^d)$ reduces to the narrow convergence). The second assumption is slightly stronger than λ -convexity along geodesics (corresponding to the case when either $\mu = \sigma_0$ or $\mu = \sigma_1$), but it happens that the conditions imposed on the internal, potential and interaction energy functionals to

ensure convexity along geodesics, ensure (5.61) as well. The same phenomenon occurs for $-W_2^2(\cdot, \nu)$, that turns out to satisfy (5.61) with $\lambda = -1$.

In [9] (see in particular Theorem 11.3.2 therein) one can find more general results where one imposes only compactness with respect to the narrow topology of $\mathcal{P}(\mathbb{R}^d)$ and convexity along geodesics: in this case one has to impose that both ϕ and $|\partial\phi|$ are lower semicontinuous with respect to the narrow convergence, an assumption that is fulfilled in many cases of interest. However, the proof of these convergence results is much harder, compared to the one presented here, and it involves a deep *variational* interpolation argument due to De Giorgi.

THEOREM 5.8 (Existence and approximation of gradient flows). *Let us assume that $\phi: \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ satisfy (5.1a)–(5.1d) and at least one of the conditions (5.60), (5.61) hold. Then for every $\mu_0 \in \overline{D(\phi)}$ there exists a unique solution μ_t of the gradient flow (according to Definition 5.2) satisfying the Cauchy condition*

$$\lim_{t \downarrow 0} \mu_t = \mu_0 \quad \text{in } \mathcal{P}_2(\mathbb{R}^d). \quad (5.62)$$

Moreover, for every choice of the discrete initial values M_τ^0 satisfying (5.52), the discrete solutions $\bar{M}_\tau(t)$ converge to μ_t in $\mathcal{P}_2(\mathbb{R}^d)$, uniformly in each bounded time interval.

Finally, if condition (5.61) holds with $\lambda \geq 0$ and $M_\tau^0 = \mu_0 \in D(\phi)$, for every $t = k\tau \in \mathcal{P}_\tau$ we have the a priori error estimate

$$W_2^2(\mu_t, \bar{M}_\tau(t)) \leq \tau(\phi(\mu_0) - \phi_\tau(\mu_0)) \leq \frac{\tau^2}{2} |\partial\phi|^2(\mu_0), \quad (5.63)$$

where we set

$$\phi_\tau(\mu) := \inf_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \phi(\nu) + \frac{1}{2\tau} W_2^2(\mu, \nu) = \inf_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\tau, \mu; \nu). \quad (5.64)$$

We give two separate proofs of this result, in the coercive case and in the strongly convex case. For the sake of simplicity, we also assume that $\phi \geq 0$ and $\mu_0 \in D(\phi)$; the a priori estimates needed in the more general coercive case can be found in [9].

PROOF OF THEOREM 5.8 IN THE COERCIVE CASE. *A priori estimates.* We easily have

$$\frac{\tau}{2} \frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} + \phi(M_\tau^n) \leq \phi(M_\tau^{n-1}), \quad (5.65)$$

which yields

$$\phi(M_\tau^n) \leq \phi(M_\tau^0) \quad \forall n \in \mathbb{N}, \quad \sum_{n=1}^{+\infty} \frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} \leq \frac{2\phi(M_\tau^0)}{\tau}. \quad (5.66)$$

In terms of $\bar{M}_\tau$, this means that

$$\sup_{t \geq 0} \phi(\bar{M}_\tau(t)) \leq \phi(M_\tau^0) \quad \forall \tau > 0. \quad (5.67)$$

From the last inequality of (5.66) we get, for $0 \leq m \leq n$,

$$\begin{aligned} W_2(M_\tau^n, M_\tau^m) &\leq \tau \sum_{k=m+1}^n \frac{W_2(M_\tau^k, M_\tau^{k-1})}{\tau} \\ &\leq \tau \left(\sum_{k=1}^n \frac{W_2^2(M_\tau^k, M_\tau^{k-1})}{\tau^2} \right)^{1/2} (n-m)^{1/2} \\ &\leq (2\phi(M_\tau^0))^{1/2} ((n-m)\tau)^{1/2}. \end{aligned} \quad (5.68)$$

Compactness and limit trajectory μ_t . (5.68) and (5.52) show that in each bounded interval $(0, T)$ the values $\{\phi(\bar{M}_\tau(t))\}_{\tau>0}$ are bounded and $\{\bar{M}_\tau(t)\}_{\tau>0}$ are bounded in $\mathcal{P}_2(\mathbb{R}^d)$, thus belong to a fixed compact set of $\mathcal{P}_2(\mathbb{R}^d)$ thanks to the coercivity assumption (5.60).

By connecting every pair of consecutive discrete values M_τ^{n-1}, M_τ^n with a constant speed geodesic parametrized in the interval $[t_\tau^{n-1}, t_\tau^n]$, we obtain by (5.68) a family of Lipschitz curves $\hat{M}_\tau$ satisfying

$$\begin{aligned} W_2(\hat{M}_\tau(t), \hat{M}_\tau(s)) &\leq C(t-s)^{1/2}, \\ W_2(\hat{M}_\tau(t), \bar{M}_\tau(t)) &\leq C\sqrt{\tau} \quad \forall t, s \in [0, T], \end{aligned} \quad (5.69)$$

where C is a constant independent of τ . Since the curves $\hat{M}_\tau$ are uniformly equicontinuous w.r.t. W_2 , Ascoli–Arzelà theorem yields the relative compactness of the family $\{\hat{M}_{\tau_h}\}_{h \in \mathbb{N}}$ in $C^0([0, T]; \mathcal{P}_2(\mathbb{R}^d))$ for each bounded interval $[0, T]$; we can therefore extract a vanishing sequence (τ_h) such that $\bar{M}_{\tau_h}(t) \rightarrow \mu_t$ in $\mathcal{P}_2(\mathbb{R}^d)$ for any $t \in [0, +\infty)$.

Space–time measures and construction of $\mathbf{v}$. Recall that $\mathbf{t}_\tau^n$ is the optimal transport map pushing M_τ^n to M_τ^{n-1} , and that the discrete velocity vector $\mathbf{V}_\tau^n$ is defined by $(\mathbf{i} - \mathbf{t}_\tau^n)/\tau$. Let us introduce the piecewise constant interpolants

$$\bar{\mathbf{t}}_\tau(t) := \mathbf{t}_\tau^n \quad \text{if } t \in ((n-1)\tau, n\tau]. \quad (5.70)$$

For every bounded time interval $I_T := (0, T]$, denoting by $X_T := \mathbb{R}^d \times I_T$, we can canonically identify $T^{-1}\bar{M}_\tau$ and $T^{-1}\mu$ to elements of $\mathcal{P}_2(X_T)$ simply by integrating with respect to the (normalized) Lebesgue measure $T^{-1}\mathcal{L}^1$ in I_T . Therefore $\bar{\mathbf{V}}_\tau$ is a vector field in $L^2(\bar{M}_\tau; \mathbb{R}^d)$ and (5.66) yields

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} |\bar{\mathbf{V}}_\tau(x, t)|^2 d\bar{M}_\tau(x) dt &= \int_{X_T} |\bar{\mathbf{V}}_\tau(x, t)|^2 d\bar{M}_\tau(x, t) \\ &= \tau \sum_{n=1}^{+\infty} \|\mathbf{V}_\tau^n\|_{L^2(M_\tau^n)}^2 \leq 2\phi(\mu_0). \end{aligned} \quad (5.71)$$

Hence, by Theorem 4.6, and taking into account the convergence in $\mathcal{P}_2(X_T)$ of $T^{-1}\bar{M}_\tau$ to $T^{-1}\mu$, the family $\bar{\mathbf{V}}_\tau$ has limit points as $\tau \downarrow 0$. We denote by $\mathbf{v}$ the limit (up to the extraction of a further subsequence, not relabeled) of $\mathbf{V}_{\tau_h}$.

Then, (4.46) and (5.71) give

$$\int_{X_T} |\mathbf{v}(x, t)|^2 d\mu(x, t) \leq \liminf_{h \rightarrow \infty} \int_{X_T} |\bar{\mathbf{V}}_{\tau_h}(x, t)|^2 d\bar{M}_{\tau_h}(x, t) \leq 2\phi(\mu_0). \quad (5.72)$$

The limits μ , $\mathbf{v}$ satisfy the continuity equation (5.4b). The following argument was introduced in [57]. Let us first observe that for every $\psi \in C_c^\infty(\mathbb{R}^d)$ we have

$$\begin{aligned} & \int_{\mathbb{R}^d} \psi(x) d\bar{M}_\tau(t)(x) - \int_{\mathbb{R}^d} \psi(x) d\bar{M}_\tau(t - \tau)(x) \\ &= \int_{\mathbb{R}^d} (\psi(x) - \psi(\bar{\mathbf{t}}_\tau(x, t))) d\bar{M}_\tau(t)(x) \\ &= \int_{\mathbb{R}^d} \langle \nabla \psi(x), x - \bar{\mathbf{t}}_\tau(x, t) \rangle d\bar{M}_\tau(t)(x) + \varepsilon(\tau, \psi, t) \\ &= \tau \int_{\mathbb{R}^d} \langle \nabla \psi(x), \bar{\mathbf{V}}_\tau(x) \rangle d\bar{M}_\tau(t)(x) + \varepsilon(\tau, \psi, t), \end{aligned}$$

where, for a suitable constant C_ψ depending only on the second derivatives of ψ

$$\begin{aligned} |\varepsilon(\tau, \psi, t)| &= \left| \int_{\mathbb{R}^d} (\psi(x) - \psi(\bar{\mathbf{t}}_\tau(x, t)) - \nabla \psi(x) \cdot (x - \bar{\mathbf{t}}_\tau(x, t))) d\bar{M}_\tau(t)(x) \right| \\ &\leq C_\psi \int_{\mathbb{R}^d} |x - \bar{\mathbf{t}}_\tau(x, t)|^2 d\bar{M}_\tau(t)(x) \\ &= C_\psi \tau^2 \int_{\mathbb{R}^d} |\bar{\mathbf{V}}_\tau(x, t)|^2 d\bar{M}_\tau(t)(x). \end{aligned}$$

Choosing now $\varphi \in C_c^\infty(\mathbb{R}^d \times (0, T))$, applying the estimate above with $\psi(\cdot) = \varphi(\cdot, t)$ and taking into account (5.71), we have

$$\begin{aligned} & - \int_{X_T} \partial_t \varphi(x, t) d\mu(x, t) \\ &= \lim_{h \rightarrow \infty} - \int_{X_T} \partial_t \varphi(x, t) d\bar{M}_{\tau_h}(x, t) \\ &= \lim_{h \rightarrow \infty} -\tau_h^{-1} \int_{X_T} (\varphi(x, t + \tau_h) - \varphi(x, t)) d\bar{M}_{\tau_h}(x, t) \\ &= \lim_{h \rightarrow \infty} \int_{X_T} \langle \nabla \varphi(t, x), \bar{\mathbf{V}}_{\tau_h} \rangle d\bar{M}_{\tau_h}(x, t) + \tau_h^{-1} \int_0^T \varepsilon(\tau_h, \varphi(t, \cdot), t) dt \\ &= \int_{X_T} \langle \nabla \varphi(t, x), \mathbf{v} \rangle d\mu(x, t). \end{aligned}$$

The limits $\mu, \mathbf{v}$ satisfy the equation $-\mathbf{v}_t \in \partial\phi(\mu_t)$. For $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ fixed we can use the variational characterization of the subdifferential (4.37) and (5.58) to obtain

$$\phi(\sigma) \geq \phi(\bar{M}_\tau(t)) - \int_{\mathbb{R}^d} \langle \bar{s}_\tau(t) - \mathbf{i}, \bar{\mathbf{v}}_\tau(t) \rangle d\bar{M}_\tau(t) + \frac{\lambda}{2} W_2^2(\sigma, \bar{M}_\tau(t))$$

for all $\tau > 0, t > 0$, where $\bar{s}_\tau$ is the optimal transport map between $\bar{M}_\tau(t)$ and σ . Then, we choose a nonnegative $\eta \in C_c^\infty((0, T))$ with $\int \eta dt = 1$ and integrate in time the previous inequality multiplied by $\eta(t)$ to find

$$\begin{aligned} \phi(\sigma) &\geq \int_0^T \phi(\bar{M}_\tau(t)) \eta(t) dt - \int_{X_T} \langle \bar{s}_\tau(t) - \mathbf{i}, \bar{\mathbf{v}}_\tau(t) \rangle d\bar{M}_\tau(t) \eta(t) dt \\ &\quad + \frac{\lambda}{2} \int_0^T W_2^2(\sigma, \bar{M}_\tau(t)) \eta(t) dt. \end{aligned} \quad (5.73)$$

Next, we set $\tau = \tau_h$ in (5.73) and pass to the limit as $h \rightarrow \infty$. By the lower semicontinuity of ϕ and the convergence of $\bar{M}_{\tau_h}(t)$ to μ_t , the convergence of the first and third integrals in the right-hand side is trivial. Concerning the second integrals, their passage to the limit is ensured by the time-dependent version of Lemma 4.7, see Remark 4.9. Therefore we obtain

$$\begin{aligned} \phi(\sigma) &\geq \int_0^T \phi(\mu_t) \eta(t) dt - \int_{X_T} \langle \mathbf{t}_{\mu_t}^\sigma - \mathbf{i}, \mathbf{v}_t \rangle d\mu_t \eta(t) dt \\ &\quad + \frac{\lambda}{2} \int_0^T W_2^2(\sigma, \mu_t) \eta(t) dt. \end{aligned}$$

If $\bar{t} \in (0, T)$ is a Lebesgue point for the map

$$t \mapsto \int_{\mathbb{R}^d} \langle \mathbf{t}_{\mu_t}^\sigma - \mathbf{i}, \mathbf{v}_t \rangle d\mu_t,$$

choosing a family η_i converging to $\delta_{\bar{t}}$ in the inequality above we get

$$\phi(\sigma) \geq \phi(\mu_{\bar{t}}) - \int_{\mathbb{R}^d} \langle \mathbf{t}_{\mu_{\bar{t}}}^\sigma - \mathbf{i}, \mathbf{v}_{\bar{t}} \rangle d\mu_{\bar{t}} + \frac{\lambda}{2} W_2^2(\sigma, \mu_{\bar{t}}).$$

As σ is arbitrary, (4.37) again gives that $-\mathbf{v}_{\bar{t}} \in \partial\phi(\mu_{\bar{t}})$.

In conclusion, the uniqueness of gradient flows gives that $\mu, \mathbf{v}$ do not depend on the chosen subsequence, and so there is full convergence as $\tau \downarrow 0$. Finally, a simple compactness argument based on the equi-continuity of $\bar{M}_\tau$ gives the local uniform convergence in $[0, +\infty)$. $\square$

PROOF OF THEOREM 5.8 IN THE STRONGLY CONVEX CASE. We shall only give a brief sketch of the proof (showing a rough error estimate, still sufficient to prove convergence)

in a simplified setting, by assuming that the strong convexity assumption (5.61) holds for $\lambda \geq 0$, ϕ is nonnegative, and $\mu_0, M_\tau^0 \in D(\phi)$.

As a preliminary remark, let us observe that if σ_s is defined as in (5.61), we have

$$\begin{aligned} W_2^2(\mu, \sigma_s) &= \int_{\mathbb{R}^d} |(1-s)\mathbf{t}_\mu^{\sigma_0} + s\mathbf{t}_\mu^{\sigma_1} - \mathbf{i}|^2 d\mu \\ &= \int_{\mathbb{R}^d} ((1-s)|\mathbf{t}_\mu^{\sigma_0} - \mathbf{i}|^2 + s|\mathbf{t}_\mu^{\sigma_1} - \mathbf{i}|^2 - s(1-s)|\mathbf{t}_\mu^{\sigma_0} - \mathbf{t}_\mu^{\sigma_1}|^2) d\mu \\ &= (1-s)W_2^2(\mu, \sigma_0) + sW_2^2(\mu, \sigma_1) - s(1-s) \int_{\mathbb{R}^d} |\mathbf{t}_\mu^{\sigma_0} - \mathbf{t}_\mu^{\sigma_1}|^2 d\mu \\ &\leq (1-s)W_2^2(\mu, \sigma_0) + sW_2^2(\mu, \sigma_1) - s(1-s)W_2^2(\sigma_0, \sigma_1). \end{aligned} \quad (5.74)$$

This inequality reflects a nice convexity property of the functional Φ defined in (5.53) and provides the starting point of our estimates.

A “metric variational inequality” for M_τ^n . The first step consists in writing a variational inequality for the discrete solution, analogous to (5.7): here we will use in a crucial way (5.61) and (5.74). In fact, it is easy to see that they yield the following strong convexity property for the functionals $s \mapsto \Phi(\tau, \mu; \sigma_s)$

$$\begin{aligned} \Phi(\tau, \mu; \sigma_s) &\leq (1-s)\Phi(\tau, \mu; \sigma_0) + s\Phi(\tau, \mu; \sigma_1) \\ &\quad - \frac{1}{2\tau}s(1-s)W_2^2(\sigma_0, \sigma_1). \end{aligned} \quad (5.75)$$

Starting from the minimum property (5.53) and applying (5.75) with $\mu := M_\tau^{n-1}$, $\sigma_0 := M_\tau^n$, $\sigma := \sigma_1 \in D(\phi)$, we get

$$\begin{aligned} &\Phi(\tau, M_\tau^{n-1}; M_\tau^n) \\ &\leq \Phi(\tau, M_\tau^{n-1}; \sigma_s) \\ &\leq (1-s)\Phi(\tau, M_\tau^{n-1}; M_\tau^n) + s\Phi(\tau, M_\tau^{n-1}; \sigma) - \frac{1}{2\tau}s(1-s)W_2^2(M_\tau^n, \sigma). \end{aligned}$$

The minimum condition says that the right derivative at $s = 0$ of the right-hand side is nonnegative; thus we find

$$\begin{aligned} &\Phi(\tau, M_\tau^{n-1}; \sigma) - \Phi(\tau, M_\tau^{n-1}; M_\tau^n) - \frac{1}{2\tau}W_2^2(M_\tau^n, \sigma) \geq 0 \\ &\forall \sigma \in D(\phi), \end{aligned} \quad (5.76)$$

which can also be written as

$$\begin{aligned} &\frac{1}{\tau} \left(\frac{1}{2}W_2^2(M_\tau^n, \sigma) - \frac{1}{2}W_2^2(M_\tau^{n-1}, \sigma) \right) \\ &\leq \phi(\sigma) - \phi(M_\tau^n) - \frac{1}{2\tau}W_2^2(M_\tau^n, M_\tau^{n-1}). \end{aligned} \quad (5.77)$$

A continuous formulation of (5.77). We want to write (5.77) as a true differential evolution inequality for the discrete solution $\bar{M}_\tau$, in order to compare two discrete solutions corresponding to different time steps τ , $\eta > 0$, and to try to reproduce the same comparison argument which we used in Theorem 5.5. Therefore, we set

$$\phi_\tau(t) := \text{“the linear interpolant of } \phi(M_\tau^{n-1}) \text{ and } \phi(M_\tau^n)\text{”} \quad \text{if } t \in (t_\tau^{n-1}, t_\tau^n],$$

i.e.,

$$\phi_\tau(t) := \frac{t_\tau^n - t}{\tau} \phi(M_\tau^{n-1}) + \frac{t - t_\tau^{n-1}}{\tau} \phi(M_\tau^n), \quad t \in (t_\tau^{n-1}, t_\tau^n]. \quad (5.78)$$

Analogously, for any $\sigma \in D(\phi)$ we set

$$\begin{aligned} W_\tau^2(t; \sigma) &:= \frac{t_\tau^n - t}{\tau} W_2^2(M_\tau^{n-1}, \sigma) \\ &\quad + \frac{t - t_\tau^{n-1}}{\tau} W_2^2(M_\tau^n, \sigma), \quad t \in (t_\tau^{n-1}, t_\tau^n]. \end{aligned} \quad (5.79)$$

Since

$$\frac{d}{dt} W_\tau^2(t; \sigma) = \frac{1}{\tau} (W_2^2(M_\tau^n, \sigma) - W_2^2(M_\tau^{n-1}, \sigma)), \quad t \in (t_\tau^{n-1}, t_\tau^n],$$

neglecting the last negative term, (5.77) becomes

$$\frac{d}{dt} \frac{1}{2} W_\tau^2(t; \sigma) \leq \phi(\sigma) - \phi_\tau(t) + \frac{1}{2} \mathcal{R}_\tau(t) \quad \forall t \in (0, T) \setminus \mathcal{P}_\tau, \quad (5.80)$$

where we set, for $t \in (t_\tau^{n-1}, t_\tau^n]$,

$$\frac{1}{2} \mathcal{R}_\tau(t) := \phi_\tau(t) - \phi(M_\tau^n) = \frac{t_\tau^n - t}{\tau} (\phi(M_\tau^{n-1}) - \phi(M_\tau^n)) \geq 0. \quad (5.81)$$

The comparison argument. We consider now another time step $\eta > 0$ inducing the partition $\mathcal{P}_\eta$, a corresponding discrete solution (M_η^k) , and the piecewise linear interpolating functions

$$\begin{aligned} W_{\tau, \eta}^2(t, s) &:= \frac{t_\eta^k - s}{\eta} W_\tau^2(t, M_\eta^k) + \frac{s - t_\eta^{k-1}}{\eta} W_\tau^2(t, M_\eta^{k-1}), \\ s &\in (t_\eta^{k-1}, t_\eta^k], \end{aligned} \quad (5.82)$$

observing that

$$W_{\tau, \eta}^2(t, s) = W_{\eta, \tau}^2(s, t) \quad \forall s, t \geq 0, \quad W_{\tau, \eta}^2(t_\tau^n, s_\eta^k) = W_2^2(M_\tau^n, M_\eta^k). \quad (5.83)$$

Taking a convex combination w.r.t. the variable $s \in I_\tau^k$ of (5.80) written for $\sigma := M_\eta^{k-1}$ and $\sigma := M_\eta^k$, we easily get

$$\frac{\partial}{\partial t} \frac{1}{2} W_{\tau, \eta}^2(t, s) \leq \phi_\eta(s) - \phi_\tau(t) + \frac{1}{2} \mathcal{R}_\tau(t), \quad t \in (0, +\infty) \setminus \mathcal{P}_\tau, s > 0. \quad (5.84)$$

Reversing the rôles of η and τ , and recalling (5.83), we also find

$$\frac{\partial}{\partial s} \frac{1}{2} W_{\tau, \eta}^2(t, s) \leq \phi_\tau(t) - \phi_\eta(s) + \frac{1}{2} \mathcal{R}_\eta(s), \quad t > 0, s \in (0, +\infty) \setminus \mathcal{P}_\eta. \quad (5.85)$$

Summing (5.84) and (5.85) we end up with

$$\begin{aligned} \frac{\partial}{\partial t} W_{\tau, \eta}^2(t, s) + \frac{\partial}{\partial s} W_{\tau, \eta}^2(s, t) &\leq \mathcal{R}_\tau(t) + \mathcal{R}_\eta(s), \\ t &\in (0, +\infty) \setminus \mathcal{P}_\tau, s \in (0, +\infty) \setminus \mathcal{P}_\eta. \end{aligned} \quad (5.86)$$

Choosing $s = t$ we eventually find

$$\frac{d}{dt} W_{\tau, \eta}^2(t, t) \leq \mathcal{R}_\tau(t) + \mathcal{R}_\eta(t), \quad t \in (0, \infty) \setminus (\mathcal{P}_\tau \cup \mathcal{P}_\eta), \quad (5.87)$$

and therefore, being $t \mapsto W_{\tau, \eta}^2(t, t)$ continuous,

$$W_{\tau, \eta}^2(T, T) \leq W_{\tau, \eta}^2(0, 0) + \int_0^T (\mathcal{R}_\tau(t) + \mathcal{R}_\eta(t)) dt \quad \forall T > 0. \quad (5.88)$$

Observe now that

$$\begin{aligned} \int_0^{+\infty} \mathcal{R}_\tau(t) dt &= \sum_{j=1}^{+\infty} \int_{t_\tau^{j-1}}^{t_\tau^j} \mathcal{R}_\tau(t) dt \\ &= \sum_{j=1}^{+\infty} \tau (\phi(M_\tau^{j-1}) - \phi(M_\tau^j)) \\ &\leq \tau \phi(M_\tau^0), \end{aligned} \quad (5.89)$$

so that (5.88) yields

$$W_{\tau, \eta}^2(T, T) \leq W_2^2(M_\tau^0, M_\eta^0) + \tau \phi(M_\tau^0) + \eta \phi(M_\eta^0) \quad \forall T > 0. \quad (5.90)$$

Convergence and rough error estimates. Recalling that

$$W_2^2(M_\tau^n, M_\tau^{n-1}) \leq \tau \phi(M_\tau^0), \quad W_2^2(M_\eta^k, M_\eta^{k-1}) \leq \eta \phi(M_\eta^0),$$

and that, for $t \in I_\tau^n \cap I_\eta^k$,

$$W_2^2(\bar{M}_\tau(t), \bar{M}_\eta(t)) \leq 3(W_{\tau,\eta}^2(t, t) + W_2^2(M_\tau^n, M_\tau^{n-1}) + W_2^2(M_\eta^k, M_\eta^{k-1})),$$

we get

$$\sup_{t \geq 0} W_2^2(\bar{M}_\tau(t), \bar{M}_\eta(t)) \leq 3(W_2^2(M_\tau^0, M_\eta^0) + 2\tau\phi(M_\tau^0) + 2\eta\phi(M_\eta^0)), \quad (5.91)$$

thus showing that $\tau \mapsto \bar{M}_\tau(t)$ is a Cauchy sequence in $\mathcal{P}_2(\mathbb{R}^d)$ for every $t \geq 0$. Denoting by μ_t its limit, we can pass to the limit in (5.90) as $\eta \downarrow 0$ by taking τ fixed and choosing $t \in \mathcal{P}_\tau$, thus obtaining the error estimate

$$\sup_{t \in \mathcal{P}_\tau} W_2^2(\bar{M}_\tau(t), \mu_t) \leq W_2^2(M_\tau^0, \mu_0) + \tau\phi(M_\tau^0). \quad (5.92)$$

μ_t is the gradient flow. To this aim, it suffices to check that μ_t satisfies the metric evolution variational inequality (5.7) with $\lambda = 0$ for every $\sigma \in D(\phi)$. Starting from the integrated form of (5.80) and recalling (5.89), we get, for every $0 < a < b < +\infty$,

$$\frac{1}{2}W_\tau^2(b, \sigma) - \frac{1}{2}W_\tau^2(a, \sigma) + \int_a^b \phi_\tau(t) dt \leq (b-a)\phi(\sigma) + \tau\phi(M_\tau^0). \quad (5.93)$$

Since

$$\begin{aligned} \lim_{\tau \downarrow 0} W_\tau^2(t, \sigma) &= W_2^2(\mu_t, \sigma), & \liminf_{\tau \downarrow 0} \phi_\tau(t) &\geq \phi(\mu_t), \\ \lim_{\tau \downarrow 0} \phi(M_\tau^0) &= \phi(\mu_0) < +\infty, \end{aligned}$$

we easily get

$$\begin{aligned} \frac{1}{2}W_2^2(\mu_b, \sigma) - \frac{1}{2}W_2^2(\mu_a, \sigma) + \int_a^b \phi(\mu_t) dt &\leq (b-a)\phi(\sigma) \\ \forall \sigma \in D(\phi), \end{aligned} \quad (5.94)$$

which yields (5.7). The regularization estimates of Theorem 5.7 (which depend only on the metric EVI formulation), together with (5.1c), show then that $\mu_t \in \mathcal{P}_2^a(\mathbb{R}^d)$ for $t > 0$. $\square$

5.4. Bibliographical notes

The notion of gradient flows. There are at least four possible approaches to gradient flows which can be adapted to the framework of Wasserstein spaces:

1. The “minimizing movement” approximation. We can simply consider any limit curve of the variational approximation scheme we introduced in Section 5.3, a “Generalized

minimizing movement” in the terminology suggested by De Giorgi in [36]. In the context of $\mathcal{P}_2(\mathbb{R}^d)$ this procedure has been first used in [57,70–73] and subsequently it has been applied in many different contexts, e.g., by [2,10,26,27,43,46,50–52,54,56,68,74]. It has the advantage to allow for the greatest generality of functionals ϕ , it provides a simple constructive method for proving existence of gradient flows, and it can be applied to arbitrary metric spaces, in particular to $\mathcal{P}_p(\mathbb{R}^d)$, the space of probability measures endowed with the p -Wasserstein distance.

2. *Curves of maximal slope.* We can look for absolutely continuous curves $\mu_t \in AC_{\text{loc}}^2((0, +\infty); \mathcal{P}_2(\mathbb{R}^d))$ which satisfy the differential form of the energy inequality

$$\frac{d}{dt}\phi(\mu_t) \leq -\frac{1}{2}|\mu'|^2(t) - \frac{1}{2}|\partial\phi|^2(\mu_t) \leq -|\partial\phi|(\mu_t) \cdot |\mu'| (t) \quad (5.95)$$

for $\mathcal{L}^1$ -a.e. $t \in (0, +\infty)$. This definition, introduced in a slightly different form in [37] and further developed in [9,38,64], it is still purely metric and it provides a general strategy to deduce differential properties satisfied by the limit curves of the minimizing movement scheme.

3. *The pointwise differential formulation.* It is the notion we adopted in Definition 5.2 and which requires the richest structure: since we have at our disposal a notion of tangent space and the related concepts of velocity vector field $\mathbf{v}_t$ and (sub)differential $\partial\phi(\mu_t)$, we can reproduce the simple definition of gradient flow modeled on smooth Riemannian manifold, i.e.,

$$\mathbf{v}_t \in -\partial\phi(\mu_t). \quad (5.96)$$

The a priori assumption that $\mu_t \in \mathcal{P}_2^a(\mathbb{R}^d)$ avoids subtle technical complications arising from the introduction of “plan-” (or measure valued-) subdifferentials instead of the simpler vector fields. The general theory, which also covers the case of an underlying separable Hilbert space of infinite dimension, has been presented in [9]. A different approach has been developed in [29].

4. *Systems of evolution variational inequalities (EVI).* In the case of λ -convex functionals along geodesics in $\mathcal{P}_2(\mathbb{R}^d)$, one can try to find solutions of the family of “metric” variational inequalities

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \nu) \leq \phi(\nu) - \phi(\mu_t) - \frac{\lambda}{2} W_2^2(\mu_t, \nu) \quad \forall \nu \in D(\phi). \quad (5.97)$$

This formulation can be considered as a “metric” version of B enilan [16] notion of integral solutions of contraction semigroups in Banach spaces generated by m -accretive operators; it provides the best kind of solutions, for which in particular one can prove not only uniqueness, but also various regularization effects and nice asymptotic behavior. These results are in fact completely analogous to the corresponding ones of the Hilbertian theory, thus showing that they do not strictly depend on the linearity of the underlying space.

Of course, the fact that such strong formulation always admits a solution involves the (geodesic) convexity of the functional ϕ and a crucial “curvature” properties of the distance. In [9] we discussed the role of these properties and presented new existence results in general metric spaces, extending the previous theory of [65].

Convergence of the variational approximation scheme. The variational approximation scheme is one of the basic tools for proving existence of gradient flows.

- (a) At the highest level of generality, when the functional ϕ does not satisfy any convexity or regularity assumption, one can only hope to prove the existence of a limit curve which will satisfy a sort of “relaxed” differential equation. In this case the proof relies on compactness arguments: passing to the limit in the discrete equation satisfied at each step by the approximating sequence M_t^n , one tries to write a relaxed form of the limit differential equation, assuming only narrow convergences of weak type. A possible formalization of this point of view has been discussed in [9], Theorem 11.1.6, and an application to fourth-order evolution equations is presented in [52] (see also [81] in the simpler framework of the Hilbert theory).

It may happen that under suitable closure and convexity assumptions on the sections of the subdifferential, which should be checked in each particular situation, this relaxed version coincides with the stronger one, and therefore one gets an effective solution to (5.3). Here we outlined the main points of this argument in the first proof of Theorem 5.8: in this case a final relaxation of the limit differential inclusion can be avoided, thanks to the (geodesic) convexity of the functional.

In general, this direct approach could be considered as a first basic step, which should be common to each attempt to apply the Wasserstein formalism for studying a gradient flow.

- (b) A second approach involves the *regularity of the functional* according to Definition 4.8, and still works with general distances and functionals. In this case the metric formulation of gradient flows as curves of maximal slope (see (5.95) and (5.8)) plays a crucial role.

The key ingredient, which allows to pass to the limit, is a refined discrete energy estimate (related to De Giorgi’s variational interpolation) and the lower semicontinuity of the slope, which follows from the regularity of the functional. We presented a detailed analysis of this point of view in [9].

- (c) A third approach, presented in the second proof of Theorem 5.8, can be performed only if the distance of the metric space, as in the case of $\mathcal{P}_2(\mathbb{R}^d)$, satisfies strong “curvature-like” bounds: moreover, the functional should satisfies a strong λ -convexity condition.

It extends to the Wasserstein framework previous results: the celebrated Crandall–Liggett [32] generation theorem for nonlinear contraction semigroups in Banach spaces, the optimal error estimates of [14,69,82] for gradient flows in Hilbert spaces, the convergence results of [65] in *nonpositively curved metric spaces* (we refer to [9] for a more detailed discussion).

Despite the strong convexity requirements on ϕ , which are nevertheless satisfied by all the examples of Section 4.5 in $\mathcal{P}_2(\mathbb{R}^d)$, this approach has interesting features:

- it does not require compactness assumptions of the sublevels of ϕ in $\mathcal{P}_2(\mathbb{R}^d)$: the convergence of the “minimizing movement” scheme is proved by a Cauchy-type estimate.
- it provides an explicit bound for the error between a discrete approximation and the continuous solution.
- it is well suited to study the stability of the gradient flow with respect to Γ -convergence of the generating functionals (see [9], Theorem 11.2.1).

6. Applications to evolution PDEs

In this section we present some applications of the theory developed in the previous section to some relevant PDEs. Since many approaches are obviously possible, let us briefly mention some advantages of the “Wasserstein” one:

- (a) The gradient flow formulation (5.3) suggests a general variational scheme (the minimizing movement approach, which we discussed in the previous section) to approximate the solution of (6.4a)–(6.4c): proving its convergence is interesting both from the theoretical (cf. the papers quoted at the end of the previous section) and the numerical point of view [59].
- (b) The variational scheme exhibits solutions which are a priori nonnegative, even if the equation does not satisfy any maximum principle as in the fourth-order case [52, 72].
- (c) Working in Wasserstein spaces allows for weak assumptions on the data: initial values which are general measures (as for fundamental solutions, in the linear cases) fit quite naturally in this framework.
- (d) The gradient flow structure suggests new contraction and energy estimates, which may be useful to study the asymptotic behavior of solutions to (6.4a)–(6.4c) [1, 13, 25, 29, 42, 74, 83], or to prove uniqueness under weak assumptions on the data.
- (e) The interplay with the theory of optimal transportation provides a novel point of view to get new functional inequalities with sharp constants [3, 12, 31, 40, 62, 75, 84, 85].
- (f) The variational structure provides an important tool in the study of the dependence of solutions from perturbation of the functional.
- (g) The setting in space of measures is particularly well suited when one considers evolution equations in infinite dimensions and tries to “pass to the limit” as the dimension d goes to ∞ .

First of all we mention the basic (but formal, at this level) example, which provides one of the main motivations to study this kind of gradient flows.

6.1. Gradient flows and evolutionary PDEs of diffusion type

In the space–time open cylinder $\mathbb{R}^d \times (0, +\infty)$ we look for *nonnegative solutions* $u : \mathbb{R}^d \times (0, +\infty)$ of a parabolic equation of the type

$$\partial_t u - \nabla \cdot \left(\nabla \left(\frac{\delta \mathcal{F}}{\delta u} \right) u \right) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty), \quad (6.1)$$

where

$$\frac{\delta \mathcal{F}(u)}{\delta u} = -\nabla \cdot F_p(x, u, \nabla u) + F_z(x, u, \nabla u). \quad (6.2)$$

This is the first variation of a typical integral functional as in (4.59)

$$\mathcal{F}(u) = \int_{\mathbb{R}^d} F(x, u(x), \nabla u(x)) \, dx \quad (6.3)$$

associated to a (smooth) Lagrangian $F = F(x, z, p) : \mathbb{R}^d \times [0, +\infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$.

Observe that (6.1) has the following structure:

$$\partial_t u + \nabla \cdot (\mathbf{v}u) = 0 \quad (\text{continuity equation}), \quad (6.4a)$$

$$\mathbf{v}u = u \nabla \psi \quad (\text{gradient condition}), \quad (6.4b)$$

$$\psi = -\frac{\delta \mathcal{F}(u)}{\delta u} \quad (\text{nonlinear relation}). \quad (6.4c)$$

In the case when F depends only on $z = u$ then we have

$$\frac{\delta \mathcal{F}(u)}{\delta u} = F_z(u), \quad u \nabla F_z(x, u) = \nabla L_F(u), \quad L_F(z) := z F'(z) - F(z). \quad (6.5)$$

Since we look for *nonnegative solutions* having (constant, by (6.4a), normalized) *finite mass*

$$u(x, t) \geq 0, \quad \int_{\mathbb{R}^d} u(x, t) \, dx = 1 \quad \forall t \geq 0, \quad (6.6)$$

and *finite quadratic momentum*

$$\int_{\mathbb{R}^d} |x|^2 u(x, t) \, dx < +\infty \quad \forall t \geq 0, \quad (6.7)$$

recalling Example 4.5.1, we can

$$\text{identify } u \text{ with the measures } \mu_t := u(\cdot, t) \mathcal{L}^d, \quad (6.8)$$

and we consider $\mathcal{F}$ as a functional defined in $\mathcal{P}_2(\mathbb{R}^d)$. Then any smooth positive function u is a solution of the system (6.4a)–(6.4c) if and only if μ is a solution in $\mathcal{P}_2(\mathbb{R}^d)$ of the gradient flow equation (5.3) for the functional $\mathcal{F}$.

Observe that (6.4a) coincides with (5.4b), the gradient constraint (6.4b) corresponds to the tangent condition $\mathbf{v}_t \in \text{Tan}_{\mu_t} \mathcal{P}_2(\mathbb{R}^d)$ of (5.4a), and the nonlinear coupling $\psi = -\delta \mathcal{F}(u)/\delta u$ is equivalent to the differential inclusion $\mathbf{v}_t \in -\partial \mathcal{F}(\mu_t)$ of (5.4c).

At this level of generality the equivalence between the system (6.4a)–(6.4c) and the evolution equation (5.3) is known only for smooth solution (which, by the way, may not exist); nevertheless, the point of view of gradient flow in the Wasserstein spaces, which was introduced by Otto in a series of pioneering and enlightening papers [57,71,73,74], still presents some interesting features, whose role should be discussed in each concrete case.

6.1.1. Changing the reference measure. In many situations the choice of the Lebesgue measure $\mathcal{L}^d$ as a reference measure, thus inducing the identification (6.8), looks quite natural; nevertheless there are some interesting cases where a different measure γ plays a crucial role (see, e.g., the examples of Section 4.5.4 and Section 6.3) and it may happen that an evolution PDE takes a simpler form by an appropriate choice of γ .

From the Wasserstein point of view, an integral functional ϕ inducing the gradient flow is defined on measures μ , but its *explicit* form *depends* on the reference γ , so that different PDEs involving the density of μ w.r.t. γ could arise from the same functional.

Let us suppose, e.g., that ϕ takes the integral form

$$\phi(\mu) = \mathcal{F}_\gamma(\rho) = \int_{\mathbb{R}^d} \tilde{F}(x, \rho(x), \nabla \rho(x)) d\gamma(x) \quad \text{if } \mu = \rho\gamma, \quad (6.9)$$

where γ is a probability measure induced by the (smooth) potential V , i.e.,

$$\gamma := e^{-V} \mathcal{L}^d \in \mathcal{P}_2(\mathbb{R}^d). \quad (6.10)$$

Since

$$u = \frac{d\mu}{d\mathcal{L}^d} = e^{-V} \rho \quad \text{and} \quad \nabla \rho = e^V (u \nabla V + \nabla u), \quad (6.11)$$

the integrand $\tilde{F}(x, \tilde{z}, \tilde{p})$ of (6.9) is related to the integrand F of the representation (6.3) by the relation

$$\begin{aligned} \tilde{z} &:= e^{V(x)} z, & \tilde{p} &:= e^{V(x)} (z \nabla V(x) + p) \\ F(x, z, p) &= e^{-V(x)} \tilde{F}(x, \tilde{z}, \tilde{p}) \\ &= e^{-V(x)} \tilde{F}(x, e^{V(x)} z, e^{V(x)} (z \nabla V(x) + p)). \end{aligned} \quad (6.12)$$

In this case it could be better to write the solution of the gradient flow μ_t generated by ϕ in terms of the density

$$\rho_t := \frac{d\mu_t}{d\gamma} = e^V \frac{d\mu_t}{d\mathcal{L}^d}, \quad (6.13)$$

and to use the differential operators associated with γ

$$\nabla_\gamma \rho := e^V \nabla (e^{-V} \rho) = \nabla \rho - \rho \nabla V, \quad (6.14a)$$

$$\nabla_\gamma \cdot \xi := e^V \nabla \cdot (e^{-V} \xi) = \nabla \cdot \xi - \xi \cdot \nabla V, \quad (6.14b)$$

which satisfy the “integration by parts formulae” with respect to the measure γ

$$\begin{aligned} \int_{\mathbb{R}^d} \xi \cdot \nabla \zeta \, d\gamma &= - \int_{\mathbb{R}^d} \zeta \nabla \gamma \cdot \xi \, d\gamma, \\ \int_{\mathbb{R}^d} \nabla \cdot \xi \, \zeta \, d\gamma &= - \int_{\mathbb{R}^d} \nabla \gamma \cdot \xi \, d\gamma, \end{aligned} \quad (6.15)$$

when $\zeta \in C_c^\infty(\mathbb{R}^d)$, $\xi \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$. The system (6.4a)–(6.4c) preserves the same structure and takes the form

$$\partial_t \rho + \nabla \gamma \cdot (\mathbf{v} \rho) = 0 \quad (\text{continuity equation}), \quad (6.16a)$$

$$\mathbf{v} \rho = \rho \nabla \psi \quad (\text{gradient condition}), \quad (6.16b)$$

$$\psi = - \frac{\delta \mathcal{F}_\gamma(\rho)}{\delta \rho} \quad (\text{nonlinear relation}), \quad (6.16c)$$

where

$$\frac{\delta \mathcal{F}_\gamma(\rho)}{\delta \rho} := - \nabla \gamma \cdot \tilde{F}_{\tilde{\rho}}(x, \rho, \nabla \rho) + \tilde{F}_{\tilde{z}}(x, \rho, \nabla \rho). \quad (6.17)$$

For, (6.16a) (resp. (6.16b)) can be transformed into (6.4a) (resp. (6.4b)), simply by multiplying the equation by e^{-V} and recalling (6.14b). The equivalence of (6.16c) and (6.4c) follows by a direct computation starting from (6.12): by (6.11) we get (with the obvious convention to evaluate F in $(x, u, \nabla u)$ and $\tilde{F}$ in $(x, \rho, \nabla \rho)$)

$$\begin{aligned} \frac{\delta \mathcal{F}(u)}{\delta u} &= F_z - \nabla \cdot F_p = \tilde{F}_{\tilde{z}} + \nabla V \cdot \tilde{F}_{\tilde{p}} - \nabla \cdot \tilde{F}_{\tilde{p}} \\ &= \tilde{F}_{\tilde{z}} - \nabla \gamma \cdot \tilde{F}_{\tilde{p}} = \frac{\delta \mathcal{F}_\gamma(\rho)}{\delta \rho}. \end{aligned}$$

REMARK 6.1 (Equations in bounded sets and Neumann boundary conditions). The possibility to change the reference measure is also useful to study evolution equations in a bounded open set $\Omega \subset \mathbb{R}^d$: they correspond to a measure γ whose support is included in $\overline{\Omega}$, e.g.

$$\gamma := \mathcal{L}^d|_{\Omega}.$$

Observe that in any case the family of time-dependent measures $\mu_t = u_t \mathcal{L}^d|_{\Omega}$, which solves of the gradient flow equation according to Definition 5.2, still satisfies the continuity equation (5.4b) in $\mathbb{R}^d \times (0, +\infty)$. This can be seen as a weak formulation of the continuity equation for u_t in $\Omega \times (0, +\infty)$ with Neumann boundary conditions on $\partial\Omega \times (0, +\infty)$:

$$\begin{aligned} \partial_t u_t + \nabla \cdot (\mathbf{v}_t u_t) &= 0 \quad \text{in } \Omega \times (0, +\infty), \\ u_t \mathbf{v}_t \cdot \mathbf{n} &= 0 \quad \text{on } \partial\Omega \times (0, +\infty). \end{aligned} \quad (6.18)$$

6.2. The linear transport equation for λ -convex potentials

Let $V : \mathbb{R}^d \rightarrow (-\infty, +\infty]$ be a proper, l.s.c. and λ -convex potential. We are looking for curves $t \mapsto \mu_t \in \mathcal{P}_2(\mathbb{R}^d)$ which solve the evolution equation

$$\frac{\partial}{\partial t} \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0, \quad \text{with } -\mathbf{v}_t(x) \in \partial V(x) \text{ for } \mu_t\text{-a.e. } x \in \mathbb{R}^d, \quad (6.19)$$

which is the gradient flow in $\mathcal{P}_2(\mathbb{R}^d)$ of the potential energy functional discussed in Example 3.4,

$$\mathcal{V}(\mu) := \int_{\mathbb{R}^d} V(x) d\mu(x). \quad (6.20)$$

If V is differentiable, (6.19) can also be written as

$$\frac{\partial}{\partial t} \mu_t = \nabla \cdot (\nabla V \mu_t) \quad \text{in the distribution sense.} \quad (6.21)$$

In the statement of the following theorem we denote by T the λ -contractive semigroup on $\overline{D(V)} \subset \mathbb{R}^d$ induced by the differential inclusion

$$\frac{d}{dt} T_t(x) \in -\partial V(T_t(x)), \quad T_0(x) = x \quad \forall x \in \overline{D(V)}. \quad (6.22)$$

Recall also that, according to Brezis theorem, $\frac{d}{dt} T_t(x)$ equals $-\partial^\circ V(T_t(x))$ at each point $t > 0$ of differentiability.

THEOREM 6.2. *For every $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ with $\text{supp } \mu_0 \subset \overline{D(V)}$, there exists a unique solution $(\mu_t, \mathbf{v})$ of (6.19) satisfying*

$$\lim_{t \downarrow 0} \mu_t = \mu_0, \quad \int_{\mathbb{R}^d} |\mathbf{v}_t(x)|^2 d\mu_t(x) \in L^1_{\text{loc}}(0, +\infty); \quad (6.23)$$

this solution is the gradient flow of $\mathcal{V}$ in the sense of the EVI formulation (6.19) and of the energy identity (5.9) of Theorem 5.3. In particular it induces a λ -contractive semigroup on $\{\mu \in \mathcal{P}_2(\mathbb{R}^d) : \text{supp}(\mu) \subset \overline{D(V)}\}$ and it exhibits the regularizing effect and the asymptotic behavior as in Theorem 5.7.

Moreover, for every $t > 0$ we have the representation formulas:

$$\mu_t = (T_t)_\# \mu_0, \quad \mathbf{v}_t(x) = -\partial^\circ V(x) \quad \text{for } \mu_t\text{-a.e. } x \in \mathbb{R}^d. \quad (6.24)$$

PROOF. Proposition 3.5 shows that the functional $\mathcal{V}$ satisfies (5.1a), (5.1b), (5.1d); it is also easy to check that (5.61) holds. On the other hand, $\mathcal{V}$ does not satisfy (5.1c), thus our simplified existence results cannot be directly applied. Nevertheless, the more general theory of [9] covers also this case and yields the present result.

In any case, the solution to (6.19) can also be directly constructed by the representation formula (6.24). It is immediate to check directly that if we choose μ_0 of the type

$$\mu_0 := \sum_{k=1}^K \alpha_k \delta_{x_k}, \quad \alpha_k \geq 0, \quad \sum_{k=1}^K \alpha_k = 1, \quad x_k \in D(V), \quad (6.25)$$

then

$$\mu_t = \sum_{k=1}^K \alpha_k \delta_{T_t(x_k)} = (T_t)_\# \mu_0 \quad (6.26)$$

solves (6.19) (see also Section 2.5, where the connection between characteristics and solutions of the continuity equation is studied in detail), whereas (6.23) follows by the energy identity

$$\int_a^b |\partial^\circ V(T_t(x))|^2 dt + \phi(T_b(x)) = \phi(T_a(x)) \quad \forall x \in \overline{D(V)}.$$

Arguing as in the proof of Theorem 2.21 we also get for every $\sigma \in D(\mathcal{V})$ and every $\gamma \in \Gamma_0(\mu_t, \sigma)$

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \sigma) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \mathbf{v}_t(x), x - y \rangle d\gamma(x, y) \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(V(y) - V(x) - \frac{\lambda}{2} |x - y|^2 \right) d\gamma(x, y) \\ &= \mathcal{V}(\sigma) - \mathcal{V}(\mu_t) - \frac{\lambda}{2} W_2^2(\mu_t, \sigma) \end{aligned} \quad (6.27)$$

at any t where $s \mapsto W_2(\mu_s, \sigma)$ and all $s \mapsto T_s(x_i)$ are differentiable. The measures $\mu_t = (T_t)_\# \mu_0$ thus solves the EVI formulation (5.7) of the gradient flow for every initial datum μ_0 which is a convex combination of Dirac masses in $D(V)$. A standard approximation argument via (5.17) and Theorem 5.7 yields the same result for $\mu_t = (T_t)_\# \mu_0$ and every admissible initial measure $\mu_0 \in \overline{D(V)}$: for, being $\text{supp } \mu_0 \subset \overline{D(V)}$, we can find a sequence $(\nu_n) \subset D(\mathcal{V})$ of convex combination of Dirac masses

$$\nu_n := \sum_{k=1}^{K_n} \alpha_{n,k} \delta_{x_{n,k}}, \quad \alpha_{n,k} \geq 0, \quad \sum_{k=1}^{K_n} \alpha_{n,k} = 1, \quad x_{n,k} \in D(V), \quad (6.28)$$

such that $\nu_n \rightarrow \mu_0$ in $\mathcal{P}_2(\mathbb{R}^d)$. □

6.3. Kolmogorov–Fokker–Planck equation

The aim of this section is to present a systematic study of the “Wasserstein” approach to Kolmogorov–Fokker–Planck (KFP in the following) equation, which was firstly proposed by Jordan, Kinderlehrer and Otto [57].

From this point of view, this equation is the gradient flow of the relative entropy functional discussed in Section 3.3; when the involved potential V is λ -convex, we have at our disposal all the tools to develop a self-contained variational theory for the generation of a λ -contracting semigroup in $\mathcal{P}_2(\mathbb{R}^d)$ with nice regularizing properties, independently of the growth of V (for other kind of estimates we refer to [34] and the references therein).

The particular “linear” structure of the subdifferential of the entropy yields the linearity of the semigroup. Under quite general assumptions, which can be applied to more general situations, the construction of a family of kernels and of general representation formulae is particularly easy in the Wasserstein framework, as well as the extension of the semigroup to L^p -spaces with respect to the invariant measure $\gamma := e^{-V} \mathcal{L}^d$. The λ -contractivity in $\mathcal{P}_2(\mathbb{R}^d)$ and the regularizing effect of the Wasserstein construction are also crucial to derive the Feller property for the KFP semigroup. We also show the equivalence with the more usual approach by Dirichlet forms in $L^2(\gamma)$.

Even if the theory presented here is finite-dimensional, we tried to develop sufficiently general arguments which could be extended to an infinite-dimensional setting, taking also account of the more general theory available in [9]. It would be interesting to compare this point of view with other well-established approaches (see, e.g., [18,35]).

6.3.1. Relative entropy and Fisher information. Let us consider

$$\begin{aligned} &\text{a l.s.c. } \lambda\text{-convex potential } V : \mathbb{R}^d \rightarrow (-\infty, +\infty] \\ &\text{with } \Omega := \text{Int}(D(V)) \neq \emptyset; \end{aligned} \tag{6.29}$$

for the sake of simplicity, we assume that the reference measure induced by the potential V is a probability measure with finite quadratic moment, i.e.,

$$\gamma := e^{-V} \mathcal{L}^d \in \mathcal{P}_2(\mathbb{R}^d). \tag{6.30}$$

This condition, up to a renormalization, is always satisfied if, e.g., $\lambda > 0$. Observe that the density e^{-V} of γ with respect to $\mathcal{L}^d$ is 0 outside $\bar{\Omega} = \overline{D(V)}$. We adopt the convention to write a measure $\mu \in \mathcal{P}_2^a(\mathbb{R}^d)$ supported in $\bar{\Omega}$ as

$$\mu = u \mathcal{L}^d|_{\Omega} = \rho \gamma, \quad u = e^{-V} \rho; \tag{6.31}$$

the *relative entropy* (see Section 3.3) of μ w.r.t. γ is defined as

$$\mathcal{H}(\mu|\gamma) = \int_{\Omega} \rho \log \rho \, d\gamma = \int_{\Omega} u(\log u + V) \, dx, \tag{6.32}$$

whereas the *relative Fisher information* is defined as

$$\mathcal{I}(\mu|\gamma) := \int_{\Omega} \left| \frac{\nabla \rho}{\rho} \right|^2 d\mu = \int_{\Omega} \frac{|\nabla \rho|^2}{\rho} d\gamma = \int_{\Omega} \frac{|\nabla u + u \nabla V|^2}{u} dx \quad (6.33)$$

whenever $u, \rho \in W_{\text{loc}}^{1,1}(\Omega)$ (recall that V is locally Lipschitz in Ω); as usual, we set $\mathcal{H}(\mu|\gamma) = +\infty$ if μ is not absolutely continuous, and $\mathcal{I}(\mu|\gamma) = +\infty$ if $\rho \notin W_{\text{loc}}^{1,1}(\Omega)$.

Let us collect in the following proposition the main properties of these two functionals, we already discussed in Sections 3–5.

PROPOSITION 6.3 (Entropy and Fisher information). *Let V, γ be as in (6.29) and (6.30).*

(i) λ -convexity of the relative entropy. *The functional $\mu \mapsto \mathcal{H}(\mu|\gamma)$ is λ -displacement convex and it also satisfies the strong convexity assumption (5.61).*

(ii) Subdifferential and slope of the entropy. *A measure $\mu = \rho \gamma = u \mathcal{L}^d|_{\Omega}$ belongs to $D(\partial \mathcal{H}) = D(|\partial \mathcal{H}|)$ iff $\mathcal{I}(\mu|\gamma) < +\infty$, i.e.,*

$$\rho, u \in W_{\text{loc}}^{1,1}(\Omega) \quad \text{and} \quad \frac{\nabla \rho}{\rho} = \frac{\nabla u}{u} + \nabla V \in L^2(\mu; \mathbb{R}^d); \quad (6.34)$$

in this case

$$\xi = \partial^\circ \mathcal{H}(\mu|\gamma) \iff \xi = \frac{\nabla \rho}{\rho} \in L^2(\mu; \mathbb{R}^d), \quad (6.35)$$

so that

$$\mathcal{I}(\mu|\gamma) = \int_{\Omega} |\xi|^2 d\mu = |\partial \mathcal{H}|^2(\mu). \quad (6.36)$$

(iii) Variational inequality for the logarithmic gradient. *If $\mathcal{I}(\mu|\gamma) < +\infty$, the logarithmic gradient $\xi = \nabla \rho / \rho$ satisfies*

$$\begin{aligned} \int_{\Omega} \left((\mathbf{t}_{\mu}^{\sigma} - x) \cdot \xi + \frac{\lambda}{2} |\mathbf{t}_{\mu}^{\sigma} - x|^2 \right) d\mu &\leq \mathcal{H}(\sigma|\gamma) - \mathcal{H}(\mu|\gamma) \\ \forall \sigma &\in \mathcal{P}_2^a(\mathbb{R}^d). \end{aligned} \quad (6.37)$$

(iv) Log-Sobolev inequality. *If $\lambda > 0$ then*

$$\mathcal{H}(\mu|\gamma) \leq \frac{1}{2\lambda} \mathcal{I}(\mu|\gamma) \quad \forall \mu \in \mathcal{P}_2^a(\mathbb{R}^d). \quad (6.38)$$

(v) Derivative of the entropy along curves. *Let $\mu : t \in [0, T] \mapsto \mu_t = \rho_t \gamma \in \mathcal{P}_2(\mathbb{R}^d)$ be a continuous family of measures satisfying the continuity equation*

$$\partial_t \mu + \nabla \cdot (\mathbf{v} \mu) = 0 \quad \text{in } \mathcal{D}'(\mathbb{R}^d \times (0, T)) \quad (6.39)$$

for a Borel vector field $\mathbf{v}$ with

$$\int_0^T \int_{\Omega} |\mathbf{v}_t|^2 d\mu_t dt < +\infty, \quad \int_0^T \mathcal{I}(\mu_t | \gamma) dt < +\infty. \quad (6.40)$$

Then the map $t \mapsto \mathcal{H}(\mu_t | \gamma)$ is absolutely continuous in $[0, T]$ and for $\mathcal{L}^1$ -a.e. $t \in (0, T)$ its derivative is

$$\frac{d}{dt} \mathcal{H}(\gamma | \mu_t) = \int_{\Omega} \mathbf{v}_t \cdot \frac{\nabla \rho_t}{\rho_t} d\mu_t = \int_{\Omega} \mathbf{v}_t \cdot \nabla \rho_t d\gamma. \quad (6.41)$$

PROOF. (i) follows from Propositions 3.5 and 3.11. The generalized convexity property (5.61) follows by analogous arguments (see [9], Proposition 9.3.9).

(ii) and (iii) have been proved in Theorem 4.21 and (4.37).

(iv) follows from (5.22a).

(v) follows from the general chain rule (4.55). $\square$

6.3.2. Wasserstein formulation of the Kolmogorov–Fokker–Planck equation. Under the same assumption (6.29), (6.30) of the previous section, and recalling the differential operators of (6.14a) and (6.14b), let us introduce the Laplacian operator Δ_{γ} induced by γ ,

$$\Delta_{\gamma} \rho := \nabla_{\gamma} \cdot (\nabla \rho) = e^V \nabla \cdot (e^{-V} \nabla \rho) = \Delta \rho - \nabla \rho \cdot \nabla V, \quad (6.42)$$

and its formal adjoint (with respect to the Lebesgue measure) Fokker–Planck operator

$$\Delta_{\gamma}^* u := e^{-V} \Delta_{\gamma} (e^V u) = \nabla \cdot (\nabla u + u \nabla V). \quad (6.43)$$

Indeed, we formally have

$$\begin{aligned} e^{-V} \Delta_{\gamma} (e^V u) &= e^{-V} [\Delta (e^V u) - \nabla (e^V u) \cdot \nabla V] \\ &= e^{-V} [\nabla \cdot (e^V \nabla u + e^V u \nabla V) - \nabla (e^V u) \cdot \nabla V] \\ &= \Delta u + \nabla u \cdot \nabla V + u \Delta V = \nabla \cdot (\nabla u + u \nabla V). \end{aligned}$$

For smooth functions with compact support in Ω they satisfy

$$-\int_{\Omega} \Delta_{\gamma} \rho \zeta d\gamma = \int_{\Omega} \nabla \rho \cdot \nabla \zeta d\gamma = -\int_{\Omega} \rho \Delta_{\gamma} \zeta d\gamma, \quad (6.44)$$

$$-\int_{\Omega} \Delta_{\gamma}^* u \zeta dx = -\int_{\Omega} u \Delta_{\gamma} \zeta dx. \quad (6.45)$$

In the case of the centered Gaussian measure with variance λ^{-1} we have

$$V(x) = \frac{1}{2} \left(\lambda |x|^2 + d \log \left(\frac{2\pi}{\lambda} \right) \right), \quad \gamma = \frac{1}{(2\pi/\lambda)^{d/2}} e^{-\lambda/2 |x|^2} \mathcal{L}^d. \quad (6.46)$$

Δ_γ is the Ornstein–Uhlenbeck operator $\Delta - \lambda x \cdot \nabla$.

The general definition of gradient flow, when particularized to the relative entropy functional, reads as follows:

DEFINITION 6.4 (“Wasserstein” solutions of KFP equations). A continuous family $\mu_t = \rho_t \gamma = u_t \mathcal{L}^d|_\Omega \in C^0((0, +\infty); \mathcal{P}_2(\mathbb{R}^d))$ is a Wasserstein solution of the Kolmogorov–Fokker–Planck equation if $t \mapsto \mathcal{I}(\mu_t|\gamma)$ belongs to $L^2_{\text{loc}}(0, +\infty)$ so that for $\mathcal{L}^1$ -a.e. $t \in (0, +\infty)$

$$\rho_t, u_t \in W^{1,1}_{\text{loc}}(\Omega), \quad \xi_t = \frac{\nabla \rho_t}{\rho_t} = \frac{\nabla u_t}{u_t} + \nabla V \in L^2(\mu_t; \mathbb{R}^d), \quad (6.47)$$

and

$$\partial_t \mu_t - \nabla \cdot \left(\mu_t \frac{\nabla \rho_t}{\rho_t} \right) = 0 \quad \text{in } \mathcal{D}'(\mathbb{R}^d \times (0, +\infty)). \quad (6.48)$$

In terms of test functions (6.48) means

$$\int_0^{+\infty} \int_\Omega \left(-\partial_t \zeta + \frac{\nabla \rho_t}{\rho_t} \cdot \nabla \zeta \right) d\mu_t dt = 0 \quad \forall \zeta \in C_c^\infty(\mathbb{R}^d \times (0, +\infty)), \quad (6.49)$$

so that ρ_t satisfy the weak formulation

$$\int_0^{+\infty} \int_\Omega (-\rho_t \partial_t \zeta + \nabla \rho_t \cdot \nabla \zeta) d\gamma dt = 0 \quad \forall \zeta \in C_c^\infty(\mathbb{R}^d \times (0, +\infty)) \quad (6.50)$$

of

$$\begin{aligned} \partial_t \rho_t - \Delta_\gamma \rho_t &= 0 \quad \text{in } \Omega \times (0, +\infty), \\ e^{-V} \partial_{\mathbf{n}} \rho_t &= 0 \quad \text{on } \partial\Omega \times (0, +\infty). \end{aligned} \quad (6.51)$$

REMARK 6.5. In terms of the Lebesgue density u_t , (6.48) reads

$$\begin{aligned} \int_0^{+\infty} \int_\Omega (-u \partial_t \zeta + (\nabla u + u \nabla V) \cdot \nabla \zeta) dx dt &= 0 \\ \forall \zeta \in C_c^\infty(\mathbb{R}^d \times (0, +\infty)), \end{aligned} \quad (6.52)$$

corresponding to the Fokker–Planck equation

$$\partial_t u - \Delta_\gamma^* u = \partial_t u - \nabla \cdot (\nabla u + u \nabla V) = 0 \quad \text{in } \Omega \times (0, +\infty), \quad (6.53)$$

with homogeneous boundary conditions $(\nabla u + u \nabla V) \cdot \mathbf{n} = 0$ on $\partial\Omega \times (0, +\infty)$.

We introduce the narrowly closed and convex (both in the metric and linear sense) subset of $\mathcal{P}_2(\mathbb{R}^d)$

$$\mathcal{P}_2(\overline{\Omega}) := \{\mu \in \mathcal{P}_2(\mathbb{R}^d) : \text{supp}(\mu) \subset \overline{\Omega}\}. \quad (6.54)$$

THEOREM 6.6. *For every $\mu_0 \in \mathcal{P}_2(\overline{\Omega})$ there exists a unique Wasserstein solution $\mu_t = \rho_t \gamma = u_t \mathcal{L}^d|_{\Omega}$ of the Kolmogorov–Fokker–Planck equation (6.48) satisfying $\mu_t \rightarrow \mu_0$ in $\mathcal{P}_2(\mathbb{R}^d)$ as $t \downarrow 0$ and it coincides with the Wasserstein gradient flow generated by the functional $\phi(\mu) := \mathcal{H}(\mu|\gamma)$.*

The maps $S_t : \mu_0 \mapsto \mu_t$, $t \geq 0$, define a continuous λ -contractive semigroup in $\mathcal{P}_2(\overline{\Omega})$ which can be characterized by the system of EVI

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \sigma) + \frac{\lambda}{2} W_2^2(\mu_t, \sigma) \leq \mathcal{H}(\sigma|\gamma) - \mathcal{H}(\mu_t|\gamma) \quad \forall \sigma \in \mathcal{P}_2(\overline{\Omega}). \quad (6.55)$$

It exhibits the regularizing effect

$$\mathcal{H}(\mu_t|\gamma) < +\infty, \quad \mathcal{I}(\mu_t|\gamma) < +\infty \quad \forall t > 0, \quad (6.56)$$

with, for $\lambda \geq 0$,

$$\mathcal{H}(\mu_t|\gamma) \leq \frac{1}{2t} W_2^2(\mu_t, \gamma), \quad \mathcal{I}(\mu_t|\gamma) \leq \frac{1}{t^2} W_2^2(\mu_t, \gamma). \quad (6.57)$$

The map $t \mapsto e^{2\lambda t} \mathcal{I}(\mu_t|\gamma)$ is nonincreasing and it satisfies the energy identity

$$\mathcal{H}(\mu_b|\gamma) + \int_a^b \mathcal{I}(\mu_t|\gamma) dt = \mathcal{H}(\mu_a|\gamma) \quad \forall 0 \leq a \leq b \leq +\infty. \quad (6.58)$$

When $\lambda > 0$ the asymptotic behavior of μ_t as $t_0 \leq t \rightarrow +\infty$ is governed by

$$\begin{aligned} W_2(\mu_t, \gamma) &\leq e^{-\lambda(t-t_0)} W_2(\mu_{t_0}, \gamma), & \mathcal{H}(\mu_t|\gamma) &\leq e^{-2\lambda(t-t_0)} \mathcal{H}(\mu_{t_0}|\gamma), \\ \mathcal{I}(\mu_t|\gamma) &\leq e^{-2\lambda(t-t_0)} \mathcal{I}(\mu_{t_0}|\gamma). \end{aligned} \quad (6.59)$$

Moreover, for every $t > 0$ (and also for $t = 0$, provided $\mathcal{I}(\mu_0|\gamma) < +\infty$)

$$\begin{aligned} \exists \lim_{h \downarrow 0} \frac{\mathbf{t}_{\mu_t}^{\mu_{t+h}} - \mathbf{i}}{h} &= \frac{\nabla \rho_t}{\rho_t} \quad \text{in } L^2(\mu_t; \mathbb{R}^d), \\ \exists \lim_{h \downarrow 0} \frac{\mathcal{H}(\mu_{t+h}|\gamma) - \mathcal{H}(\mu_t|\gamma)}{h} &= \mathcal{I}(\mu_t|\gamma). \end{aligned} \quad (6.60)$$

PROOF. Is is not difficult to check that

$$\overline{D(\phi)} = \mathcal{P}_2(\overline{\Omega}). \quad (6.61)$$

In fact, $D(\phi)$ contains all the measures of the type

$$\mu_{x_0, \rho} := \frac{1}{\gamma(B_\rho(x_0))} \chi_{B_\rho(x_0)} \cdot \gamma \quad \text{with } B_\rho(x_0) \subset \subset \Omega,$$

and their convex combinations, so that

$$\sum_i \alpha_i \delta_{x_i} \in \overline{D(\phi)} \quad \text{if } x_i \in \Omega, \alpha_i \geq 0, \sum_i \alpha_i = 1.$$

Since the subset of all the finite convex combinations of δ -measures concentrated in Ω is dense in $\mathcal{P}_2(\overline{\Omega})$, we get (6.61).

By Proposition 6.3 the relative entropy functional $\mu \mapsto \mathcal{H}(\mu|\gamma)$ satisfies all the assumptions of Theorem 5.3, Theorem 5.7, and Theorem 5.8 (in the strongly convex case). Theorem 6.6 is a simple transposition of the results of Section 5, taking also into account the particular form of the subdifferential of $\mathcal{H}$ expressed by (6.35) and the fact that γ is the unique minimum of $\mathcal{H}$ with $\mathcal{H}(\gamma|\gamma) = 0$. $\square$

We conclude this section by briefly discussing some further properties of the semigroup constructed by Theorem 6.6. We first introduce the “transition probabilities” $\nu_{x,t} = \vartheta_{x,t} \gamma$

$$\nu_{x,t} := S_t[\delta_x] \quad \text{with densities } \vartheta_{x,t} := \frac{d\nu_{x,t}}{d\gamma} \in L^1(\gamma) \quad \forall x \in \overline{\Omega}, t > 0. \quad (6.62)$$

Besicovitch differentiation theorem and the narrow continuity of $x \mapsto S_t[\delta_x]$ give that the explicit formula

$$\vartheta_{x,t}(y) := \limsup_{r \downarrow 0} \frac{\nu_{x,t}(B_r(y))}{\gamma(B_r(y))}$$

provides us with a pointwise definition of the densities $\vartheta_{x,t}$ satisfying

$$\text{for every } t > 0 \text{ the map } (x, y) \in \overline{\Omega} \times \overline{\Omega} \rightarrow \vartheta_{x,t}(y) \text{ is Borel.} \quad (6.63)$$

THEOREM 6.7 (The associated Markovian semigroup). *Let $(S_t)_{t \geq 0}$ be the semigroup constructed in the previous Theorem 6.6 and let us consider the set of densities*

$$B_\gamma := \{\rho \in L^1(\gamma) : \rho\gamma \in \mathcal{P}_2(\mathbb{R}^d)\}. \quad (6.64)$$

Extension to a contraction semigroup in $L^p(\gamma)$. There exists a unique strongly continuous semigroup of linear contraction operators $(\mathcal{S}_t)_{t \geq 0}$ in $L^1(\gamma)$ such that

$$\mathcal{S}_t[\rho_0] = \rho_t \iff S_t[\rho_0\gamma] = \rho_t\gamma \quad \forall \rho_0 \in B_\gamma. \quad (6.65)$$

For every $p \in [1, +\infty]$ $\mathcal{S}_t$ is a continuous (only weakly continuous, if $p = +\infty$) contraction semigroup in $L^p(\gamma)$*

$$\|\mathcal{S}_t[\rho]\|_{L^p(\gamma)} \leq \|\rho\|_{L^p(\gamma)} \quad \forall \rho \in L^p(\gamma), \quad (6.66)$$

it is order preserving

$$\rho_0 \leq \rho_1 \implies \mathfrak{J}_t[\rho_0] \leq \mathfrak{J}_t[\rho_1], \quad (6.67)$$

and regularizing, since

$$\mathfrak{J}_t(L^\infty(\gamma)) \subset C_b(\overline{\Omega}) \quad \forall t > 0. \quad (6.68)$$

Moreover,

$$\begin{aligned} \mathfrak{J}_t(\text{Lip}(\Omega)) &\subset \text{Lip}(\Omega) & \forall t > 0, \\ \text{Lip}(\mathfrak{J}_t[\rho]; \Omega) &\leq e^{-\lambda t} \text{Lip}(\rho, \Omega) & \forall \rho \in \text{Lip}(\Omega). \end{aligned} \quad (6.69)$$

Representation formula. The semigroups S_t , $\mathfrak{J}_t$ admit the representation formulas

$$S_t[\mu] = \rho_t \gamma \quad \text{with } \rho_t(x) = \int_{\mathbb{R}^d} \vartheta_{y,t}(x) d\mu(y) \quad \gamma\text{-a.e.} \quad (6.70)$$

$$\mathfrak{J}_t[\rho_0] = \rho_t \quad \text{with } \rho_t(x) = \int_{\Omega} \vartheta_{y,t}(x) \rho_0(y) d\gamma(y) \quad \gamma\text{-a.e.} \quad (6.71)$$

Dirichlet form. $\mathfrak{J}_t$ coincides in $L^2(\gamma)$ with the (analytic) semigroup $\widetilde{\mathfrak{J}}_t$ associated to the symmetric Dirichlet form with domain

$$W_\gamma^{1,2}(\Omega) := \{\rho \in W_{\text{loc}}^{1,2}(\Omega) : \rho \in L^2(\gamma), \nabla \rho \in L^2(\gamma; \mathbb{R}^d)\} \subset L^2(\gamma), \quad (6.72)$$

$$a_\gamma(\rho, \eta) := \int_{\Omega} \nabla \rho \cdot \nabla \eta d\gamma \quad \forall \rho, \eta \in W_\gamma^{1,2}(\Omega). \quad (6.73)$$

In particular, if $\rho_0 \in L^2(\gamma)$ then the solution $\rho_t = \mathfrak{J}_t[\rho_0]$ satisfies

$$\rho \in L_{\text{loc}}^2([0, +\infty); W_\gamma^{1,2}(\Omega)) \cap C^0([0, +\infty); L^2(\gamma)) \quad (6.74)$$

and

$$\begin{aligned} \int_0^{+\infty} (-(\rho, \partial_t \eta)_{L^2(\gamma)} + a_\gamma(\rho, \eta)) dt &= 0 \\ \forall \eta \in C_c^1((0, +\infty); W_\gamma^{1,2}(\Omega)). \end{aligned} \quad (6.75)$$

Symmetry of the transition densities. For every $t > 0$ the transition densities $\vartheta_{x,t}$ satisfy

$$\vartheta_{x,t}(y) = \vartheta_{y,t}(x) \quad \text{for } \gamma \times \gamma\text{-a.e. } (x, y) \in \Omega \times \Omega, \quad (6.76)$$

so that the “adjoint” representation formula holds

$$\mathfrak{J}_t[\rho_0] = \rho_t \quad \text{with } \rho_t(x) = \int_{\Omega} \vartheta_{x,t}(y) \rho_0(y) d\gamma(y), \quad (6.77)$$

which provides the continuous representative of ρ_t when $\rho_0 \in L^\infty(\gamma)$.

PROOF. Most of the results stated in the theorem are a direct consequence of the “linearity” of the semigroup $\mathcal{S}_t$ and of its regularizing effect; therefore, we postpone their proof to the next section, where we will discuss from a general point of view the construction of a Markov semigroup starting from a “linear” Wasserstein semigroup.

Here we only consider the last two properties, establishing the link with the “Dirichlet form” approach. Let us first observe that $W_\gamma^{1,2}(\Omega)$ is dense in $L_\gamma^2(\Omega)$ and it is an Hilbert space with the norm

$$\|\rho\|_{W_\gamma^{1,2}(\Omega)}^2 := \|\rho\|_{L^2(\gamma)}^2 + a_\gamma(\rho, \rho) = \int_\Omega (|\rho|^2 + |\nabla \rho|^2) d\gamma. \quad (6.78)$$

In fact, this is equivalent to the lower semicontinuity property of a_γ with respect to convergence in L_γ^2

$$\begin{cases} \rho_n \in W_\gamma^{1,2}(\Omega), & \rho_n \rightarrow \rho \text{ in } L^2(\gamma), \\ \sup_n a_\gamma(\rho_n, \rho_n) \leq C \end{cases} \implies \rho \in W_\gamma^{1,2}(\Omega), \quad a_\gamma(\rho, \rho) \leq C. \quad (6.79)$$

Formulation (6.75) is stronger than the Wasserstein one as ρ is supposed to be in $L_{\text{loc}}^2([0, +\infty); W_\gamma^{1,2}(\Omega))$; whenever this extra regularity holds, then more general test functions in $W_\gamma^{1,2}(\Omega)$ are allowed in (6.50), since it is not difficult to check that $C_c^\infty(\mathbb{R}^d)$ functions are dense in $W_\gamma^{1,2}(\Omega)$; it is then possible to recover (6.75) directly from (6.50).

The main idea is then to prove that a Wasserstein solution starting from $\mu_0 := \rho_0 \gamma$ with $\rho_0 \in L^2(\gamma)$ satisfies the energy estimate (in fact an identity)

$$2 \int_0^T \int_\Omega |\nabla \rho_t|^2 d\gamma dt + \int_\Omega |\rho_T|^2 d\gamma \leq \int_\Omega |\rho_0|^2 d\gamma \quad \forall T > 0, \quad (6.80)$$

by evaluating the time derivative of the $L^2(\gamma)$ -norm of ρ along the solution of the gradient flow.

For, we need a preliminary regularization and we consider the family of real convex superlinear functions $F_k : [0, +\infty) \rightarrow [0, +\infty)$ (depending on $k > 0$)

$$F_k(\rho) := \begin{cases} \rho^2 & \text{if } \rho \leq k, \\ k\rho(1 - \log k + \log \rho) & \text{if } \rho \geq k, \end{cases} \quad (6.81)$$

which satisfy

$$0 \leq F_k(\rho) \leq c_k + k\rho \log \rho, \quad F_k(\rho) \uparrow \rho^2 \text{ as } k \uparrow +\infty \quad \forall \rho \geq 0. \quad (6.82)$$

F_k induces the relative energy functional

$$\mathcal{F}_k(\mu|\gamma) := \int_\Omega F_k(\rho) d\gamma. \quad (6.83)$$

A direct calculations shows that F_k satisfies (3.22) and

$$L_{F_k}(\rho) = \begin{cases} \rho^2 & \text{if } \rho \leq k, \\ k\rho & \text{if } \rho \geq k, \end{cases} \quad (6.84)$$

so that for a measure $\mu = \rho\gamma$

$$\xi = \partial^\circ \mathcal{F}_k(\mu|\gamma) \iff \rho \in W_\gamma^{1,1}(\Omega), \quad \xi = \frac{\nabla L_{F_k}(\rho)}{\rho} \in L^2(\gamma; \mathbb{R}^d). \quad (6.85)$$

Being L_F Lipschitz, the chain rule for Sobolev functions $\rho \in W_\gamma^{1,1}(\Omega)$ yields

$$\nabla L_{F_k}(\rho) = \begin{cases} 2\rho \nabla \rho & \text{in } \Omega \cap \{x: \rho(x) \leq k\}, \\ k \nabla \rho & \text{in } \Omega \cap \{x: \rho(x) > k\}. \end{cases} \quad (6.86)$$

If $\mathcal{I}(\mu|\gamma) < +\infty$ then $\mu \in D(\partial \mathcal{F}_k)$ since

$$\int_\Omega \left| \frac{\nabla L_{F_k}(\rho)}{\rho} \right|^2 \rho \, d\gamma \leq 4k^2 \mathcal{I}(\mu|\gamma) < +\infty. \quad (6.87)$$

If $\mu_0 = \rho_0\gamma$, $\rho_0 \in L^2(\gamma)$ then $\mathcal{F}_k(\mu_0|\gamma) < +\infty$, $\mathcal{H}(\mu_0|\gamma) < +\infty$, and the chain rule (4.55) yields

$$\mathcal{F}_k(\mu_t|\gamma) + \int_0^T \int_\Omega \frac{\nabla L_{F_k}(\rho_t) \cdot \nabla \rho_t}{\rho_t} \, d\gamma \, dt = \mathcal{F}_k(\rho_0) \leq \int_\Omega |\rho_0|^2 \, d\gamma < +\infty. \quad (6.88)$$

By (6.86),

$$\int_\Omega \frac{\nabla F_k(\rho_t) \cdot \nabla \rho_t}{\rho_t} \, d\gamma \geq 2 \int_{\Omega \cap \{\rho_t \leq k\}} |\nabla \rho_t|^2 \, d\gamma,$$

so that the monotone convergence theorem yields (6.80).

Let us now check the last statement of Theorem 6.7. (6.75) and the regularity (6.74) yield that for every $\eta \in W_\gamma^{1,2}(\Omega)$ the map

$$\begin{aligned} t \mapsto \int_\Omega \mathfrak{J}_t[\rho] \eta \, d\gamma & \text{ is absolutely continuous, with} \\ \frac{d}{dt} \int_\Omega \mathfrak{J}_t[\rho] \eta \, d\gamma + a_\gamma(\mathfrak{J}_t[\rho], \eta) & = 0. \end{aligned} \quad (6.89)$$

By integrating (6.89) and choosing initial data $\rho, \eta \in W_\gamma^{1,2}$, being a_γ a symmetric form it is immediate to check that $\mathfrak{J}_t$ is self-adjoint in $L^2(\gamma)$ and we have

$$\int_\Omega \rho \mathfrak{J}_t[\eta] \, d\gamma = \int_\Omega \mathfrak{J}_t[\rho] \eta \, d\gamma \quad \forall \rho, \eta \in L^2(\gamma). \quad (6.90)$$

For every bounded nonnegative $\rho, \eta \in L^\infty(\gamma)$, (6.71) yields

$$\begin{aligned} & \int_{\Omega} \rho(x) \left(\int_{\Omega} \vartheta_{y,t}(x) \eta(y) d\gamma(y) \right) d\gamma(x) \\ &= \int_{\Omega} \left(\int_{\Omega} \vartheta_{x,t}(y) \rho(x) d\gamma(x) \right) \eta(y) d\gamma(y). \end{aligned} \quad (6.91)$$

By (6.63) and Fubini's theorem, we get (6.76).

Finally, the fact that $\mathcal{J}$ is a continuous semigroup in $L^1(\gamma)$ follows directly from the estimate (6.80): being $\mathcal{J}_t$ nonexpansive, it is sufficient to check that $\mathcal{J}_t[\rho] \rightarrow \rho$ strongly in $L^1(\gamma)$ as $t \downarrow 0$ on the dense subset $L^2(\gamma)$. The uniform bound of (6.80) provides both the weak and the strong convergence of $\mathcal{J}_t[\rho_0]$ to ρ_0 in $L^2(\gamma)$ as $t \downarrow 0$. $\square$

REMARK 6.8 (Dirichlet forms and analytic Markovian semigroups). Since the variational solution of (6.75) is unique (by Lions' theorem on variational evolution equations in a Hilbert triplet, see, e.g., [23]), in the proof of Theorem 6.7 we do not really need the converse implication showing that solutions of (6.75) are Wasserstein solutions of (6.50) with

$$\int_a^b \mathcal{I}(\rho_t \gamma | \gamma) dt < +\infty \quad \forall 0 < a < b < +\infty. \quad (6.92)$$

Nevertheless, we briefly mention how one can pass from (6.75) to the Wasserstein formulation; the main point is to show that the relative Fisher information is locally integrable in $(0, +\infty)$.

Let us first recall that for every $\rho_0 \in L^2(\gamma)$ Lions' theorem provides a unique solution

$$\begin{aligned} \rho &\in L^2_{\text{loc}}([0, +\infty); W^{1,2}_\gamma(\Omega)) \cap H^1_{\text{loc}}([0, +\infty); (W^{1,2}_\gamma(\Omega))') \\ &\subset C^0([0, +\infty); L^2(\gamma)) \end{aligned}$$

solving (6.75) or, equivalently,

$$\begin{aligned} & \frac{d}{dt} \int_{\Omega} \rho \eta d\gamma + \int_{\Omega} \nabla \rho \cdot \nabla \eta d\gamma = 0 \\ & \forall \eta \in W^{1,2}_\gamma(\Omega), \mathcal{L}^1\text{-a.e. in } (0, +\infty), \end{aligned} \quad (6.93)$$

and such that $\lim_{t \downarrow 0} \rho_t = \rho_0$ strongly in $L^2(\gamma)$. Moreover, ρ_t satisfies the energy identity

$$\int_0^T \int_{\Omega} |\nabla \rho_t|^2 d\gamma dt + \frac{1}{2} \int_{\Omega} |\rho_T|^2 d\gamma = \frac{1}{2} \int_{\Omega} |\rho_0|^2 d\gamma, \quad (6.94)$$

and since a_γ is symmetric the map $\mathcal{J}_t: \rho_0 \mapsto \rho_t$ is a contraction *analytic* semigroup in $L^2(\gamma)$. In particular, ρ enjoys the nicer property $\rho \in C^\infty((0, +\infty); W^{1,2}_\gamma(\Omega))$.

Moreover, a standard truncation argument in Sobolev space yields

$$a_\gamma(\rho^+ \wedge 1, \rho^+ \wedge 1) \leq a_\gamma(\rho, \rho) \quad \forall \rho \in W_\gamma^{1,2}(\Omega), \quad (6.95)$$

so that a_γ is a *closed and symmetric Dirichlet form* in $L^2(\gamma)$ (see, e.g., [63]); in particular

$$\mathfrak{J}_t(c) = c \quad \forall c \in \mathbb{R}; \quad \rho_0 \leq \rho_1 \implies \mathfrak{J}_t(\rho_0) \leq \mathfrak{J}_t(\rho_1). \quad (6.96)$$

In order to check the equivalence with the Wasserstein formulation, we observe that for every initial datum $\rho_0 \in L^\infty(\gamma)$ with $\rho_0(x) \geq r > 0$ for γ -a.e. $x \in \Omega$, the unique solution ρ_t of (6.93) still satisfies the lower bound $\rho_t \geq r$ by (6.96); moreover, (6.94) yields

$$\int_0^{+\infty} \mathcal{I}(\rho_t | \gamma) dt \leq r^{-1} \int_0^{+\infty} \int_\Omega |\nabla \rho_t|^2 d\gamma dt < +\infty, \quad (6.97)$$

so that, by Theorem 5.3, the measures $\mu_t = \rho_t \gamma$ provide the unique Wasserstein solution of (6.50) (since γ is a finite measure, $C_c^\infty(\mathbb{R}^d)$ is a subset of $W_\gamma^{1,2}(\Omega)$). Therefore the semigroups $\mathfrak{S}$ and $\tilde{\mathfrak{S}}$ coincide on $L^2(\gamma)$ -densities bounded away from 0: a simple density argument shows that they coincide on $L^2(\gamma)$.

REMARK 6.9. The measures $(\nu_{x,t})_{t \geq 0}$ are a Markovian semigroup of kernels associated with $(\mathcal{S}_t)_{t \geq 0}$ ([63], Section II-4).

6.3.3. The construction of the Markovian semigroup. Among general λ -contracting semigroups in $\mathcal{P}_2(\mathbb{R}^d)$, the Kolmogorov–Fokker–Planck equation enjoys several other interesting features, due to its linearity. As we will see in the next lemma, this is a direct consequence of the following “linearity condition”

$$\begin{aligned} & \begin{cases} \xi_i = \partial^\circ \phi(\mu_i), & \alpha_i \geq 0, & \alpha_1 + \alpha_2 = 1, \\ \xi(\alpha_1 \mu_1 + \alpha_2 \mu_2) = \alpha_1 \xi_1 \mu_1 + \alpha_2 \xi_2 \mu_2 \end{cases} \\ \implies & \xi \in \partial \phi(\alpha_1 \mu_1 + \alpha_2 \mu_2) \end{aligned} \quad (6.98)$$

satisfied by the Wasserstein subdifferential of $\phi(\mu) := \mathcal{H}(\mu | \gamma)$.

The aim of this section is to show how easily one can deduce contraction and regularizing estimates starting from a “linear” Wasserstein semigroup; in particular, the construction of the fundamental solutions is particularly simple. It should not be too difficult to extend the following results to infinite dimensional underlying spaces, taking into account that the existence and the uniqueness of the gradient flow of the relative entropy functional extend to this context (see [9]).

LEMMA 6.10 (Linearity of the gradient flow). *Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ be a functional satisfying (5.1a)–(5.1d) and let S_t be the λ -contractive semigroup generated by its*

gradient flow on $\overline{D(\phi)}$ as in Theorem 5.7. If ϕ satisfies (6.98), then the semigroup S_t satisfies the “linearity” property

$$\begin{aligned} S_t[\alpha_1\mu_1 + \alpha_2\mu_2] &= \alpha_1 S_t[\mu_1] + \alpha_2 S_t[\mu_2] \\ \forall \mu_1, \mu_2 \in \overline{D(\phi)}, \alpha_1, \alpha_2 \geq 0, \alpha_1 + \alpha_2 &= 1. \end{aligned} \quad (6.99)$$

PROOF. Take two initial data $\mu_1, \mu_2 \in \overline{D(\phi)}$ and set $\mu_{i,t} := S_t[\mu_i]$, $\mathbf{v}_{i,t} = -\partial^\circ \phi(\mu_{i,t})$ their velocity vector fields, $\mu_t = \alpha_1\mu_{1,t} + \alpha_2\mu_{2,t}$, and define the vector field $\mathbf{v}_t$ so that

$$\mathbf{v}_t \mu_t := \alpha_1 \mathbf{v}_{1,t} \mu_{1,t} + \alpha_2 \mathbf{v}_{2,t} \mu_{2,t}. \quad (6.100)$$

Assuming $\alpha_i > 0$ and introducing the densities

$$\rho_{i,t} := \frac{d\mu_{i,t}}{d\mu_t}, \quad \text{so that} \quad \mathbf{v}_t = \alpha_1 \rho_{1,t} \mathbf{v}_{1,t} + \alpha_2 \rho_{2,t} \mathbf{v}_{2,t}, \quad \alpha_1 \rho_{1,t} + \alpha_2 \rho_{2,t} = 1,$$

it is easy to check that, for every $t > 0$,

$$\begin{aligned} \int_{\mathbb{R}^d} |\mathbf{v}_t|^2 d\mu_t &= \int_{\mathbb{R}^d} |\alpha_1 \rho_{1,t} \mathbf{v}_{1,t} + \alpha_2 \rho_{2,t} \mathbf{v}_{2,t}|^2 d\mu_t \\ &\leq \alpha_1 \int_{\mathbb{R}^d} |\mathbf{v}_{1,t}|^2 \rho_{1,t} d\mu_t + \alpha_2 \int_{\mathbb{R}^d} |\mathbf{v}_{2,t}|^2 \rho_{2,t} d\mu_t \\ &= \alpha_1 \int_{\mathbb{R}^d} |\mathbf{v}_{1,t}|^2 d\mu_{1,t} + \alpha_2 \int_{\mathbb{R}^d} |\mathbf{v}_{2,t}|^2 d\mu_{2,t}. \end{aligned} \quad (6.101)$$

It follows that the map $t \mapsto \|\mathbf{v}_t\|_{L^2(\mu_t; \mathbb{R}^d)}$ belongs to $L^2_{\text{loc}}(0, +\infty)$ and, by linearity, μ_t satisfies the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty). \quad (6.102)$$

Since $\mathbf{v}_t \in \partial \phi(\mu_t)$ by (6.98), μ_t is the unique gradient flow with initial datum $\alpha_1\mu_1 + \alpha_2\mu_2$, that is $\mu_t = S_t(\alpha_1\mu_1 + \alpha_2\mu_2)$. $\square$

Let γ be a nonnegative Borel measure on $\mathbb{R}^d$, with support D , and let B_γ be defined as in (6.64).

THEOREM 6.11. *For $t \geq 0$, let $S_t: \mathcal{P}_2(D) \rightarrow \mathcal{P}_2(D)$ be satisfying the following assumptions:*

$$S_t \text{ is a continuous } \lambda\text{-contracting semigroup.} \quad (6.103a)$$

$$S_t[\mu] \ll \gamma \quad \forall \mu \in \mathcal{P}_2(D), t > 0. \quad (6.103b)$$

$$\begin{aligned} S_t[\alpha\mu + \beta\nu] &= \alpha S_t[\mu] + \beta S_t[\nu] \\ \forall \mu, \nu \in \mathcal{P}_2(D), \alpha, \beta \geq 0, \alpha + \beta &= 1. \end{aligned} \quad (6.103c)$$

Then the following properties hold.

Extension to $L^1(\gamma)$. There exists a unique narrowly continuous semigroup (denoted by $\mathfrak{S}_t$) of bounded linear operators on $L^1(\gamma)$ such that

$$S_t[\rho\gamma] = \mathfrak{S}_t[\rho]\gamma \quad \forall \rho \in B_\gamma. \quad (6.104)$$

Contraction and order preserving properties. $\mathfrak{S}_t$ is in fact a contraction and order preserving semigroup, i.e.,

$$\|\mathfrak{S}_t[\rho]\|_{L^1(\gamma)} \leq \|\rho\|_{L^1(\gamma)}, \quad \rho_1 \leq \rho_2 \implies \mathfrak{S}_t[\rho_1] \leq \mathfrak{S}_t[\rho_2]. \quad (6.105)$$

Representation formula. Denoting by $\nu_{t,x} = \vartheta_{t,x}\gamma$ the “transition probabilities”

$$\nu_{x,t} := S_t[\delta_x], \quad \text{with densities } \vartheta_{x,t} := \frac{d\nu_{x,t}}{d\gamma} \in L^1(\gamma) \quad \forall x \in D, t > 0, \quad (6.106)$$

the semigroup S_t admits the representation formula

$$S_t[\mu] = \rho_t\gamma \quad \text{with } \rho_t(x) = \int_{\mathbb{R}^d} \vartheta_{y,t}(x) d\mu(y) \text{ for } \gamma\text{-a.e. } x \in D. \quad (6.107)$$

Invariant measure and Markov property. If

$$\gamma \in \mathcal{P}_2(\mathbb{R}^d) \text{ is an invariant measure, i.e., } S_t[\gamma] = \gamma \quad \forall t \geq 0, \quad (6.108)$$

then

$$\mathfrak{S}_t(L^p(\gamma)) \subset L^p(\gamma) \quad \forall p \in [1, +\infty] \quad (6.109)$$

and the restriction of $\mathfrak{S}_t$ to $L^p(\gamma)$ is a continuous (weakly* continuous if $p = \infty$) contraction semigroup.

PROOF. Let us first extend S by homogeneity to the cone $\mathcal{M}_2(D)$ of nonnegative finite measures with finite second moment

$$\mathcal{M}_2(D) := \{\lambda\mu: \mu \in \mathcal{P}_2(D), \lambda \geq 0\} \quad (6.110)$$

simply by setting

$$S_t[\lambda\mu] = \lambda S_t[\mu] \quad \forall \mu \in \mathcal{P}_2(D), \lambda \geq 0. \quad (6.111)$$

It is easy to check that this extension preserves properties (6.103a) and (6.103b) and, moreover, (6.103c) holds for every couple of nonnegative coefficients α, β :

$$S_t[\alpha\mu + \beta\nu] = \alpha S_t[\mu] + \beta S_t[\nu] \quad \forall \mu, \nu \in \mathcal{M}_2(D), \alpha, \beta \geq 0. \quad (6.112)$$

The uniqueness of $\mathcal{S}_t$ is then immediate: if $\rho\gamma \in \mathcal{M}_2(D)$ then by (6.104) and (6.103b)

$$\mathcal{S}_t[\rho] = \frac{S_t[\rho\gamma]}{\gamma}. \quad (6.113)$$

Being $\mathcal{S}_t$ continuous, it is sufficient to determine it on the set

$$C_\gamma := \left\{ \rho \in L^1(\gamma): \int |x|^2 |\rho(x)| d\gamma(x) < +\infty \right\}, \quad (6.114)$$

which is clearly dense in $L^1(\gamma)$; since each $\rho \in C_\gamma$ can be decomposed as

$$\rho = \rho_+ - \rho_-, \quad \text{where } \rho_+\gamma, \rho_-\gamma \in \mathcal{M}_2(D), \quad (6.115)$$

$\mathcal{S}_t[\rho]$ should be equal to the difference between $\mathcal{S}_t[\rho_+]$ and $\mathcal{S}_t[\rho_-]$. Let us check that this representation is independent of the particular decomposition: if ρ'_+, ρ'_- is another admissible couple as in (6.115), then $\rho_+ + \rho'_- = \rho'_+ + \rho_-$ and therefore

$$\mathcal{S}_t[\rho_+] + \mathcal{S}_t[\rho'_-] = \mathcal{S}_t[\rho_+ + \rho'_-] = \mathcal{S}_t[\rho'_+ + \rho_-] = \mathcal{S}_t[\rho'_+] + \mathcal{S}_t[\rho_-],$$

showing that

$$\mathcal{S}_t[\rho_+] - \mathcal{S}_t[\rho_-] = \mathcal{S}_t[\rho'_+] - \mathcal{S}_t[\rho'_-].$$

Choosing, in particular, $\rho_+ := \max[\rho, 0]$ and $\rho_- := -\min[\rho, 0]$ we get the bound

$$\begin{aligned} \|\mathcal{S}_t[\rho]\|_{L^1(\gamma)} &\leq \|\mathcal{S}_t[\rho_+]\|_{L^1(\gamma)} + \|\mathcal{S}_t[\rho_-]\|_{L^1(\gamma)} \\ &= \|\rho_+\|_{L^1(\gamma)} + \|\rho_-\|_{L^1(\gamma)} = \|\rho\|_{L^1(\gamma)}, \end{aligned} \quad (6.116)$$

which shows that $\mathcal{S}_t$ is nonexpansive. Therefore, it can also be uniquely extended to a nonexpansive linear operator on $L^1(\gamma)$.

From the narrow continuity of $x \mapsto S_t[\delta_x]$ we also get

$$\text{the map } x \mapsto \int_D \varphi(y) \vartheta_{x,t}(y) d\gamma(y) \text{ is continuous } \forall \varphi \in C_b^0(\mathbb{R}^d). \quad (6.117)$$

In order to prove the representation formula (6.107) we observe that for every initial measure $\nu = \sum_i \alpha_i \delta_{x_i} \in \mathcal{P}_2(D)$ and every $\varphi \in C_b^0(D)$, $\nu_t = S_t[\nu]$ satisfies

$$\begin{aligned} \int_D \varphi(y) d\nu_t(y) &= \sum_i \alpha_i \int_D \varphi(y) \vartheta_{x_i,t}(y) d\gamma(y) \\ &= \int_D \left(\int_D \varphi(y) \vartheta_{x,t}(y) d\gamma(y) \right) d\nu(x). \end{aligned} \quad (6.118)$$

Therefore, by approximating in $\mathcal{P}_2(\mathbb{R}^d)$ an arbitrary measure $\mu \in \mathcal{P}_2(D)$ by a sequence of concentrated measures $\nu^k = \sum_i \alpha_i^k \delta_{x_i^k}$, since $S_t[\nu^k] \rightarrow S_t[\mu] = \mu_t = \rho_t \gamma$ in $\mathcal{P}_2(\mathbb{R}^d)$, (6.117) yields

$$\int_D \varphi(y) \rho_t(y) d\gamma(y) = \int_D \left(\int_D \varphi(y) \vartheta_{x,t}(y) d\gamma(y) \right) d\mu(x), \quad (6.119)$$

and therefore (6.107) follows by Fubini's theorem.

Finally, if γ is an invariant measure, then $\mathcal{J}_t[1] = 1$; the order preserving property shows that $\|\mathcal{J}_t[\rho]\|_{L^\infty(\gamma)} \leq \|\rho\|_{L^\infty(\gamma)}$. By interpolation, the same property holds for every space $L^p(\gamma)$. $\square$

In order to study the adjoint semigroup $\mathcal{J}^*$ of $\mathcal{J}$ we further suppose that

$$\sup_{y \in D \cap B_r(x_0)} \int_D \varphi(\vartheta_{y,t}(x)) d\gamma(x) < +\infty \quad \forall x_0 \in D, t, r > 0, \quad (6.120a)$$

for some continuous convex function $\varphi: [0, +\infty) \rightarrow [0, +\infty)$ with more than linear growth at infinity, and

$$\limsup_{t \downarrow 0} \int_D \varphi(\mathcal{J}_t[\rho](x)) d\gamma(x) < +\infty \quad \forall \rho \in B_\gamma \cap L^\infty(\gamma). \quad (6.120b)$$

In the case of the KFP semigroup we have seen that these properties hold with $\varphi(z) = z \ln z$. For every function $\zeta \in L^\infty(\gamma)$ we can thus define

$$\zeta_t(x) = \mathcal{J}_t^*[\zeta](x) := \int_{\mathbb{R}^d} \vartheta_{x,t}(y) \zeta(y) d\gamma(y). \quad (6.121)$$

The next result show that $\mathcal{J}_t^*$ is the adjoint semigroup of $\mathcal{J}_t$ and it exhibits the Feller regularizing property.

THEOREM 6.12 (The adjoint semigroup). *Under the same assumption of the previous theorem and (6.120a), (6.120b), $\mathcal{J}_t$ is a strongly continuous semigroup in $L^1(\gamma)$ and the maps $\mathcal{J}_t^*$ defined by (6.121) are the weakly*-continuous, nonexpansive, adjoint semigroup on $L^\infty(\gamma)$ induced by $\mathcal{J}_t$, i.e. they satisfy*

$$\int_{\mathbb{R}^d} \mathcal{J}_t^*[\zeta] \rho d\gamma = \int_{\mathbb{R}^d} \zeta \mathcal{J}_t[\rho] d\gamma \quad \forall \rho \in L^1(\gamma), \zeta \in L^\infty(\gamma). \quad (6.122)$$

Moreover, for every $t > 0$,

$$\mathcal{J}_t^*(L^\infty(\gamma)) \subset C_b^0(D), \quad \mathcal{J}_t^*(\text{Lip}(D)) \subset \text{Lip}(D), \quad (6.123)$$

$$\text{Lip}(\mathcal{J}_t[\rho]; D) \leq e^{-\lambda t} \text{Lip}(\rho, D) \quad \forall \rho \in \text{Lip}(D). \quad (6.124)$$

PROOF. We already know that $\mathcal{S}_t$ is a narrowly continuous contraction semigroup in $L^1(\gamma)$. For linear semigroups, strong continuity is equivalent to weak continuity [77]; therefore, being $B_\gamma \cap L^\infty(\gamma)$ a dense subset in (the positive cone of) $L^1(\gamma)$, it is sufficient to check that

$$\mathcal{S}_t[\rho_0] \rightharpoonup \rho_0 \quad \text{weakly in } L^1(\gamma) \quad \forall \rho_0 \in B_\gamma \cap L^\infty(\gamma). \quad (6.125)$$

Condition (6.125) follows then directly from the narrow continuity of the map $t \mapsto \mathcal{S}_t[\rho_0]$ and its weak compactness in $L^1(\gamma)$ given by the uniform bound (6.120b).

Let us denote by $\mathcal{S}_t^*$ the adjoint semigroup, defined as in (6.122), and by ζ_t the image of $\zeta \in L^\infty(\gamma)$ by $\mathcal{S}_t^*$; we introduce the measures

$$\gamma_{x_0}^r := \frac{1}{\gamma(B_r(x_0))} \chi_{B_r(x_0)} \gamma \in \mathcal{P}_2(\mathbb{R}^d) \quad \forall x_0 \in D = \text{supp}(\gamma), r > 0, \quad (6.126)$$

satisfying

$$\gamma_{x_0}^r \rightarrow \delta_{x_0} \quad \text{in } \mathcal{P}_2(\mathbb{R}^d) \quad \text{as } r \downarrow 0, \quad \forall x_0 \in D. \quad (6.127)$$

Let us check that the functions

$$\vartheta_{x_0,t}^r := \mathcal{S}_t \left[\frac{\chi_{B_r(x_0)}}{\gamma(B_r(x_0))} \right] = \frac{dS_t[\gamma_{x_0}^r]}{d\gamma}, \quad \vartheta_{x_0,t}^r(x) = \int_D \vartheta_{y,t}(x) d\gamma_{x_0}^r(y)$$

satisfy

$$\vartheta_{x_0,t}^r \rightharpoonup \vartheta_{x_0,t} \quad \text{weakly in } L^1(\gamma) \quad \text{as } r \downarrow 0. \quad (6.128)$$

For, narrow convergence is provided by (6.127) and the continuity of S_t in $\mathcal{P}_2(\mathbb{R}^d)$, whereas weak- $L^1(\gamma)$ compactness (when $r \in (0, r_0]$) is provided by (6.107), Jensen inequality and (6.120a) since

$$\begin{aligned} \int_D \varphi(\vartheta_{x_0,t}^r(x)) d\gamma(x) &= \int_D \varphi \left(\int_D \vartheta_{y,t}(x) d\gamma_{x_0}^r(y) \right) d\gamma(x) \\ &\leq \int_D \left(\int_D (\varphi(\vartheta_{y,t}(x)) d\gamma_{x_0}^r(y) \right) d\gamma(x) \\ &= \int_D \left(\int_D \varphi(\vartheta_{y,t}(x)) d\gamma(x) \right) d\gamma_{x_0}^r(y) \\ &\leq \sup_{y \in D \cap B_{r_0}(x_0)} \int_D \varphi(\vartheta_{y,t}(x)) d\gamma(x) < +\infty. \end{aligned}$$

It follows that for every $\zeta \in L^\infty(\gamma)$ and every $x_0 \in D, t > 0$ the limit

$$\tilde{\zeta}_t(x_0) := \lim_{r \downarrow 0} \frac{1}{\gamma(B_r(x_0))} \int_{B_r(x_0)} \zeta_t(x) d\gamma(x) \quad (6.129)$$

exists since

$$\begin{aligned} \frac{1}{\gamma(B_r(x_0))} \int_{B_r(x_0)} \zeta_t(x) \, d\gamma(x) &= \int_{\mathbb{R}^d} \mathcal{J}_t^*[\zeta] \frac{\chi_{B_r(x_0)}}{\gamma(B_r(x_0))} \, d\gamma \\ &= \int_{\mathbb{R}^d} \zeta \, \mathcal{J}_t \left[\frac{\chi_{B_r(x_0)}}{\gamma(B_r(x_0))} \right] \, d\gamma \\ &= \int_{\mathbb{R}^d} \zeta \, \vartheta_{x_0,t}^r \, d\gamma, \end{aligned}$$

and therefore

$$\tilde{\zeta}_t(x_0) = \lim_{r \downarrow 0} \int_{\mathbb{R}^d} \zeta \, \vartheta_{x_0,t}^r \, d\gamma = \int_{\mathbb{R}^d} \zeta \, \vartheta_{x_0,t} \, d\gamma = \mathcal{J}_t^*[\zeta](x_0). \quad (6.130)$$

Then, Lebesgue differentiation theorem yields $\zeta_t(x) = \mathcal{J}_t^*[\zeta](x)$ for γ -a.e. $x \in D$, thus showing that $\tilde{\mathcal{J}}^* = \mathcal{J}^*$.

From (6.120a) (providing compactness with respect to the weak L^1 topology) and the narrow continuity of $x \mapsto S_t[\delta_x]$ we obtain

$$\vartheta_{x,t} \rightharpoonup \vartheta_{x_0,t} \quad \text{weakly in } L^1(\gamma) \text{ as } x \rightarrow x_0 \, \forall t > 0, \quad (6.131)$$

and therefore $\tilde{\zeta}_t$ is the continuous representative of ζ_t ; this also shows the first inclusion of (6.123).

The second inclusion of (6.123) follows easily, since for each $\zeta \in \text{Lip}(D)$, setting $\zeta_t = \mathcal{J}_t^*(\zeta)$, for each couple of points $x, y \in D$ we have

$$\begin{aligned} |\zeta_t(x) - \zeta_t(y)| &= \left| \int_{\mathbb{R}^d} \zeta \, dS_t[\delta_x] - \int_{\mathbb{R}^d} \zeta \, dS_t[\delta_y] \right| \\ &\leq \text{Lip}(\zeta; D) W_2(S_t[\delta_x], S_t[\delta_y]) \\ &\leq \text{Lip}(\zeta; D) e^{-\lambda t} W_2(\delta_x, \delta_y) = e^{-\lambda t} \text{Lip}(\zeta; D) |x - y|. \quad \square \end{aligned}$$

6.4. Nonlinear diffusion equations

In this section we consider the case of nonlinear diffusion equations in $\mathbb{R}^d$.

Let us consider a convex differentiable function $F: [0, +\infty) \rightarrow \mathbb{R}$ which satisfies (4.70), (4.75) and (4.77): F is the density of the *internal energy functional* $\mathcal{F}$ defined in (4.69).

Setting $L_F(z) := zF'(z) - F(z)$, we are looking for *nonnegative* solution of the evolution equation

$$\partial_t u_t - \Delta(L_F(u_t)) = 0 \quad \text{in } \mathbb{R}^d \times (0, +\infty), \quad (6.132a)$$

satisfying the (normalized) mass conservation

$$u_t \in L^1(\mathbb{R}^d), \quad \int_{\mathbb{R}^d} u_t(x) \, dx = 1 \quad \forall t > 0, \quad (6.132b)$$

the finiteness of the quadratic moment

$$\int_{\mathbb{R}^d} |x|^2 u_t(x) \, dx < +\infty \quad \forall t > 0, \quad (6.132c)$$

the integrability condition $L_F(u) \in L^1_{\text{loc}}(\mathbb{R}^d \times (0, +\infty))$, and the initial Cauchy condition

$$\lim_{t \downarrow 0} u_t \cdot \mathcal{L}^d = \mu_0 \quad \text{in } \mathcal{P}_2(\mathbb{R}^d). \quad (6.132d)$$

Therefore (6.132a) has the usual distributional meaning

$$\int_0^{+\infty} \int_{\mathbb{R}^d} (-u_t \partial_t \zeta - L_F(u_t) \Delta \zeta) \, dx \, dt = 0 \quad \forall \zeta \in C_c^\infty(\mathbb{R}^d \times (0, +\infty)).$$

We can always assume possibly redefining u_t in an $\mathcal{L}^1$ -negligible set of times, that $t \mapsto u_t \mathcal{L}^d$ is narrowly continuous in $[0, +\infty)$.

THEOREM 6.13. *Suppose that F has a superlinear growth as in (4.71). Then for every $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a unique solution*

$$u \in AC^2_{\text{loc}}((0, +\infty); \mathcal{P}_2(\mathbb{R}^d))$$

of (6.132a)–(6.132d) among those satisfying

$$\begin{aligned} L_F(u) &\in L^1_{\text{loc}}((0, +\infty); W^{1,1}_{\text{loc}}(\mathbb{R}^d)), \\ \int_{\mathbb{R}^d} \frac{|\nabla L_F(u)|^2}{u} \, dx &\in L^1_{\text{loc}}(0, +\infty). \end{aligned} \quad (6.133)$$

The map $t \mapsto S_t[\mu_0] = \mu_t = u_t \mathcal{L}^d$ is the unique gradient flow in $\mathcal{P}_2(\mathbb{R}^d)$ of the functional $\mathcal{F}$ defined in (4.69), which is geodesically convex (and also satisfies (5.61) with $\lambda = 0$).

The gradient flow satisfies all properties of Theorem 5.7 for $\lambda = 0$. In particular, it is characterized by the system of EVI

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \sigma) \leq \mathcal{F}(\sigma) - \mathcal{F}(\mu_t) \quad \mathcal{L}^1\text{-a.e. } \forall \sigma \in D(\mathcal{F}), \quad (6.134)$$

it is nonexpansive

$$W_2(S_t[\mu_0], S_t[\nu_0]) \leq W_2(\mu_0, \nu_0) \quad \forall \mu_0, \nu_0 \in \mathcal{P}_2(\mathbb{R}^d) \quad (6.135)$$

and regularizing

$$\begin{aligned}
\sup_{0 < t \leq 1} t \int_{\mathbb{R}^d} F(u_t(x)) \, dx &< +\infty, \\
\sup_{0 < t \leq 1} t^2 \int_{\mathbb{R}^d} \frac{|\nabla L_F(u_t(x))|^2}{u_t(x)} \, dx &< +\infty, \\
t \mapsto \int_{\mathbb{R}^d} \frac{|\nabla L_F(u_t(x))|^2}{u_t(x)} \, dx &\text{ is nonincreasing,}
\end{aligned} \tag{6.136}$$

and for every $t > 0$,

$$\begin{aligned}
\exists \lim_{h \downarrow 0} \frac{\mathbf{t}_{\mu_t}^{\mu_{t+h}} - \mathbf{i}}{h} &= \frac{\nabla L_F(u_t)}{u_t} \quad \text{in } L^2(\mu_t; \mathbb{R}^d), \\
\exists \lim_{h \downarrow 0} \frac{\mathcal{F}(\mu_{t+h}) - \mathcal{F}(\mu_t)}{h} &= \int_{\mathbb{R}^d} \frac{|\nabla L_F(u_t(x))|^2}{u_t(x)} \, dx.
\end{aligned} \tag{6.137}$$

PROOF. The proof is a simple combination of Theorems 5.3, 5.7 and 5.8 (in the strongly convex case, see also [9], Proposition 9.3.9), and of the results of Section 4.5.3 for the functional $\mathcal{F}$, noticing that the domain of $\mathcal{F}$ is dense in $\mathcal{P}_2(\mathbb{R}^d)$. $\square$

REMARK 6.14. When F has a sublinear growth and satisfies

$$\lim_{z \rightarrow +\infty} \frac{F(z)}{z} = 0, \quad \lim_{z \rightarrow +\infty} \frac{F(z)}{z^{1-1/d}} = -\infty, \tag{6.138}$$

then it is possible to prove ([9], Theorem 10.4.8) that $\mathcal{F}$ still satisfies (5.1c) and the Wasserstein semigroup generated by $\mathcal{F}$ provides the unique solution of (6.132a) in the above precise meaning: for, even if μ_0 is not regular (e.g., a Dirac mass), the regularizing effect of the Wasserstein semigroup shows that $\mu_t := S[\mu_0](t)$ is absolutely continuous w.r.t. the Lebesgue measure $\mathcal{L}^d$ for all $t > 0$: its density u_t w.r.t. $\mathcal{L}^d$ is therefore well defined and solves (6.132a).

REMARK 6.15. Equation (6.132a) is a very classical problem: it has been studied by many authors from different points of view, which is impossible to recall in detail here.

We only mention that in the case of homogeneous Dirichlet boundary conditions in a bounded domain, Brezis showed that the equation is the gradient flow (see [22]) of the convex functional (since L_F is monotone)

$$\psi(u) := \int_{\mathbb{R}^d} G_F(u) \, dx, \quad \text{where } G_F(u) := \int_0^u L_F(r) \, dr,$$

in the space $H^{-1}(\Omega)$. We refer to the paper of Otto [74] for a detailed comparison of the two notions of solutions and for a physical justification of the interest of the Wasserstein approach.

It is also possible to prove that the differential operator $-\Delta(L_F(u))$ is m -accretive in $L^1(\mathbb{R}^d)$ and therefore it induces a (nonlinear) contraction semigroup in $L^1(\mathbb{R}^d)$. Notice that here we allow for more general initial data (an arbitrary probability measure), whereas in the H^{-1} (or L^1) formulation Dirac masses are not allowed (but see [30,78] in the fast diffusion case).

6.5. Drift diffusion equations with nonlocal terms

Let us consider, as in [28,29], a functional ϕ which is the sum of internal, potential, and interaction energy:

$$\phi(\mu) := \int_{\mathbb{R}^d} F(u) \, dx + \int_{\mathbb{R}^d} V \, d\mu + \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W \, d\mu \times \mu \quad \text{if } \mu = u \mathcal{L}^d.$$

Here F, V, W satisfy the assumptions considered in Section 4.5.7; as usual we set $\phi(\mu) = +\infty$ if $\mu \in \mathcal{P}_2(\mathbb{R}^d) \setminus \mathcal{P}_2^a(\mathbb{R}^d)$. The gradient flow of ϕ in $\mathcal{P}_2(\mathbb{R}^d)$ leads to the equation

$$\partial_t u_t - \nabla \cdot (\nabla L_F(u_t) + u_t \nabla V + u_t (\nabla W) \star u_t) = 0, \quad (6.139)$$

coupled with conditions (6.132b)–(6.132d).

THEOREM 6.16. *For every $\mu_0 \in \mathcal{P}_2(\overline{\Omega})$ there exists a unique distributional solution u_t of (6.139) among those satisfying $u_t \mathcal{L}^d \rightarrow \mu_0$ in $\mathcal{P}_2(\mathbb{R}^d)$ as $t \downarrow 0$, $L_F(u_t) \in L_{\text{loc}}^1((0, +\infty); W_{\text{loc}}^{1,1}(\Omega))$, and*

$$\left\| \frac{\nabla L_F(u_t)}{u_t} + \nabla V + (\nabla W) \star u_t \right\|_{L^2(\mu_t; \mathbb{R}^d)} \in L_{\text{loc}}^2(0, +\infty). \quad (6.140)$$

Furthermore, this solution is the unique gradient flow in $\mathcal{P}_2(\mathbb{R}^d)$ of the functional ϕ , which is λ -geodesically convex, and therefore satisfies all the properties stated in Theorem 5.7. In particular, when $\lambda > 0$ there exists a unique minimizer $\bar{\mu}$ of ϕ and the gradient flow generates a λ -contracting and regularizing semigroup which exhibits the asymptotic behavior of (5.22a)–(5.22d).

PROOF. The existence of u_t follows by Theorem 5.8 (besides (5.1) the function ϕ satisfies the strong convexity assumption (5.61), see [9], Theorem 9.3.5) and by the characterization, given in Section 4.5.7, of the (minimal) subdifferential of ϕ . The same characterization proves that any u_t as in the statement of the theorem is a gradient flow; therefore the uniqueness Theorem 5.5 can be applied. $\square$

In the limiting case $F, V = 0$, the generated semigroup loses its regularizing effect and its existence and main properties follow from the more general theory of [9]. In this way it is possible to study a model equation for the evolution of granular flows (see, e.g., [25]).

Notice that, as we did in Section 6.3, we can also consider evolution equations in convex (bounded or unbounded) domains $\Omega \subset \mathbb{R}^d$ with homogeneous Neumann boundary conditions, simply by setting $V(x) \equiv +\infty$ for $x \in \mathbb{R}^d \setminus \bar{\Omega}$.

6.6. Gradient flow of $-W^2/2$ and geodesics

For a fixed reference measure $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ let us now consider the functional $\phi(\mu) := -1/2W_2^2(\mu, \sigma)$, as in Theorem 4.20. Being ϕ (-1) -convex along generalized geodesics, we can apply Theorem 5.7 to show that ϕ generates an evolution semigroup on $\mathcal{P}_2(\mathbb{R}^d)$. The following result ([9], Theorem 11.2.10) shows that this evolution semigroup coincides with the (unique) extension of the geodesic between σ and μ_0 as long as this extension is still a minimizing geodesic.

THEOREM 6.17. *Let be given two measures $\sigma, \mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and suppose that $\gamma \in \Gamma_0(\sigma, \mu_0)$ satisfies the following property: the constant speed geodesic*

$$\gamma(s) := ((1-s)\pi^1 + s\pi^2)_\# \gamma$$

can be extended to an interval $[0, T]$, with $T > 1$. Then the formula

$$t \rightarrow \mu(t) := \gamma(e^t) \quad \text{for } 0 \leq t \leq \log(T) \quad (6.141)$$

gives the gradient flow of $\mu \mapsto -1/2W_2^2(\mu, \sigma)$ starting from μ_0 .

References

- [1] M. Agueh, *Asymptotic behavior for doubly degenerate parabolic equations*, C. R. Math. Acad. Sci. Paris **337** (2003), 331–336.
- [2] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge–Kantorovich theory*, Adv. Differential Equations **10** (2005), 309–360.
- [3] M. Agueh, N. Ghoussoub and X. Kang, *The optimal evolution of the free energy of interacting gases and its applications*, C. R. Math. Acad. Sci. Paris **337** (2003), 173–178.
- [4] G. Alberti and L. Ambrosio, *A geometrical approach to monotone functions in $\mathbb{R}^n$* , Math. Z. **230** (1999), 259–316.
- [5] A.D. Aleksandrov, *A theorem on triangles in a metric space and some of its applications*, Trudy Mat. Inst. Steklov. **38** (1951), 5–23.
- [6] L. Ambrosio, *Minimizing movements*, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5) **19** (1995), 191–246.
- [7] L. Ambrosio, *Lecture notes on optimal transport problem*, Mathematical Aspects of Evolving Interfaces, CIME Summer School in Madeira (Pt), Vol. 1812, P. Colli and J. Rodrigues, eds, Springer-Verlag (2003), 1–52.
- [8] L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Math. Monogr., Clarendon, Oxford (2000).
- [9] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Spaces of Probability Measures*, Lectures Math. ETH Zürich, Birkhäuser, Basel (2005).
- [10] L. Ambrosio, S. Lisini and G. Savaré, *Stability of Flows Associated to Gradient Vector Fields and Convergence of Iterated Transport Maps*, Manuscripta Math., to appear.

- [11] L. Ambrosio and P. Tilli, *Topics on Analysis in Metric Spaces*, Oxford Lecture Ser. Math. Appl., Vol. 25, Oxford Univ. Press, Oxford (2004).
- [12] A. Arnold and J. Dolbeault, *Refined convex Sobolev inequalities*, J. Funct. Anal. **225** (2005), 337–351.
- [13] A. Arnold, P. Markowich, G. Toscani and A. Unterreiter, *On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations*, Comm. Partial Differential Equations **26** (2001), 43–100.
- [14] C. Baiocchi, *Discretization of evolution variational inequalities*, Partial Differential Equations and the Calculus of Variations, Vol. I, F. Colombini, A. Marino, L. Modica and S. Spagnolo, eds, Birkhäuser, Boston, MA (1989), 59–92.
- [15] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem*, Numer. Math. **84** (2000), 375–393.
- [16] P. Bénéilan, *Solutions intégrales d’équations d’évolution dans un espace de Banach*, C. R. Acad. Sci. Paris **274** (1972), A47–A50.
- [17] P. Bernard and B. Buffoni, *Optimal mass transportation and Mather theory*, Journal of the European Mathematical Society, to appear.
- [18] V.I. Bogachev, *Gaussian Measures*, Math. Surveys Monogr., Vol. 62, Amer. Math. Soc., Providence, RI (1998).
- [19] F. Bouchut, F. Golse and M. Pulvirenti, *Kinetic Equations and Asymptotic Theory*, Ser. Appl. Math. (Paris), Gauthier-Villars, Éd. Sci. Méd. Elsevier, Paris (2000). Edited and with a foreword by Benoit Perthame and Laurent Desvillettes.
- [20] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), 375–417.
- [21] H. Brézis, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, Contribution to Nonlinear Functional Analysis, Proc. Sympos. Math. Res. Center, Univ. of Wisconsin Press, Madison (1971); Academic Press, New York (1971), 101–156.
- [22] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland, Amsterdam (1973). North-Holland Math. Stud., Vol. 5, Notas de Matemática (50).
- [23] H. Brezis, *Analyse fonctionnelle – Théorie et applications*, Masson, Paris (1983).
- [24] G. Buttazzo, *Semicontinuity, Relaxation and Integral Representation in the Calculus of Variations*, Pitman Res. Notes Math. Ser., Vol. 207, Longman, Harlow (1989).
- [25] E. Caglioti and C. Villani, *Homogeneous cooling states are not always good approximations to granular flows*, Arch. Ration. Mech. Anal. **163** (2002), 329–343.
- [26] E.A. Carlen and W. Gangbo, *Constrained steepest descent in the 2-Wasserstein metric*, Ann. of Math. (2) **157** (2003), 807–846.
- [27] E.A. Carlen and W. Gangbo, *Solution of a model Boltzmann equation via steepest descent in the 2-Wasserstein metric*, Arch. Ration. Mech. Anal. **172** (2004), 21–64.
- [28] J.A. Carrillo, R.J. McCann and C. Villani, *Kinetic equilibration rates for granular media and related equations: Entropy dissipation and mass transportation estimates*, Rev. Mat. Iberoamericana **19** (2003), 971–1018.
- [29] J.A. Carrillo, R.J. McCann and C. Villani, *Contractions in the 2-Wasserstein space and thermalization of granular media*, Arch. Ration. Mech. Anal. (2006).
- [30] E. Chasseigne and J.L. Vazquez, *Theory of extended solutions for fast-diffusion equations in optimal classes of data. Radiation from singularities*, Arch. Ration. Mech. Anal. **164** (2002), 133–187.
- [31] D. Cordero-Erausquin, B. Nazaret and C. Villani, *A mass-transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities*, Adv. Math. **182** (2004), 307–332.
- [32] M.G. Crandall and T.M. Liggett, *Generation of semi-groups of nonlinear transformations on general Banach spaces*, Amer. J. Math. **93** (1971), 265–298.
- [33] M.G. Crandall and A. Pazy, *Semi-groups of nonlinear contractions and dissipative sets*, J. Funct. Analysis **3** (1969), 376–418.
- [34] G. Da Prato and A. Lunardi, *Elliptic operators with unbounded drift coefficients and Neumann boundary condition*, J. Differential Equations **198** (2004), 35–52.
- [35] G. Da Prato and J. Zabczyk, *Second Order Partial Differential Equations in Hilbert Spaces*, London Math. Soc. Lecture Notes Ser., Vol. 293, Cambridge Univ. Press, Cambridge (2002).

- [36] E. De Giorgi, *New problems on minimizing movements*, Boundary Value Problems for PDE and Applications, C. Baiocchi and J. L. Lions, eds, Masson (1993), 81–98.
- [37] E. De Giorgi, A. Marino and M. Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **68** (1980), 180–187.
- [38] M. Degiovanni, A. Marino and M. Tosques, *Evolution equations with lack of convexity*, Nonlinear Anal. **9** (1985), 1401–1443.
- [39] C. Dellacherie and P.-A. Meyer, *Probabilities and Potential*, North-Holland Math. Stud., Vol. 29, North-Holland, Amsterdam (1978).
- [40] M. Del Pino, J. Dolbeault and I. Gentil, *Nonlinear diffusions, hypercontractivity and the optimal L^p -Euclidean logarithmic Sobolev inequality*, J. Math. Anal. Appl. **293** (2004), 375–388.
- [41] R.J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math. **98** (1989), 511–547.
- [42] J. Dolbeault, D. Kinderlehrer and M. Kowalczyk, *Remarks about the flashing ratchet*, Technical Report 406, Ceremade (2004).
- [43] L.C. Evans, W. Gangbo and O. Savin, *Diffeomorphisms and nonlinear heat flows*, SIAM J. Math. Anal., to appear.
- [44] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions*, Stud. Adv. Math., CRC Press, Boca Raton, FL (1992).
- [45] H. Federer, *Geometric Measure Theory*, Grundlehren Math. Wiss., Bd 153, Springer-Verlag, New York (1969).
- [46] J. Feng and M. Katsoulakis, *A Hamilton–Jacobi theory for controlled gradient flows in infinite dimensions*, Technical report (2003).
- [47] W. Gangbo, *The Monge mass transfer problem and its applications*, Monge Ampère Equation: Applications to Geometry and Optimization Deerfield Beach, FL (1997); Contemp. Math., Vol. 226, Amer. Math. Soc., Providence, RI (1999), 79–104.
- [48] W. Gangbo and R.J. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), 113–161.
- [49] R. Gardner, *The Brunn–Minkowski inequality*, Bull. Amer. Math. Soc. **39** (2002), 355–405.
- [50] L. Giacomelli and F. Otto, *Variational formulation for the lubrication approximation of the Hele–Shaw flow*, Calc. Var. Partial Differential Equations **13** (2001), 377–403.
- [51] L. Giacomelli and F. Otto, *Rigorous lubrication approximation*, Interfaces Free Bound. **5** (2003), 483–529.
- [52] U. Gianazza, G. Toscani and G. Savaré, *The Wasserstein gradient flow of the Fisher information and the Quantum Drift-Diffusion equation*, Technical report, IMATI-CNR, Pavia (2006).
- [53] M. Giaquinta and S. Hildebrandt, *Calculus of Variations I*, Grundlehren Math. Wiss., Vol. 310, Springer-Verlag, Berlin (1996).
- [54] K. Glasner, *A diffuse interface approach to Hele–Shaw flow*, Nonlinearity **16** (2003), 49–66.
- [55] C. Goffman and J. Serrin, *Sublinear functions of measures and variational integrals*, Duke Math. J. **31** (1964), 159–178.
- [56] C. Huang and R. Jordan, *Variational formulations for Vlasov–Poisson–Fokker–Planck systems*, Math. Methods Appl. Sci. **23** (2000), 803–843.
- [57] R. Jordan, D. Kinderlehrer and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM J. Math. Anal. **29** (1998), 1–17 (electronic).
- [58] J. Jost, *Nonpositive Curvature: Geometric and Analytic Aspects*, Lectures Math. ETH Zürich, Birkhäuser, Basel (1997).
- [59] D. Kinderlehrer and N.J. Walkington, *Approximation of parabolic equations using the Wasserstein metric*, M2AN Math. Model. Numer. Anal. **33** (1999), 837–852.
- [60] M. Knott and C.S. Smith, *On the optimal mapping of distributions*, J. Optim. Theory Appl. **43** (1984), 39–49.
- [61] Y. Kōmura, *Nonlinear semi-groups in Hilbert space*, J. Math. Soc. Japan **19** (1967), 493–507.
- [62] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Comm. Pure Appl. Math., to appear.
- [63] Z.-M. Ma and M. Röckner, *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*, Springer-Verlag, New York (1992).
- [64] A. Marino, C. Saccon and M. Tosques, *Curves of maximal slope and parabolic variational inequalities on nonconvex constraints*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (4) **16** (1989), 281–330.

- [65] U.F. Mayer, *Gradient flows on nonpositively curved metric spaces and harmonic maps*, Comm. Anal. Geom. **6** (1998), 199–253.
- [66] R.J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), 153–179.
- [67] R. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (2001), 589–608.
- [68] T. Mikami, *Dynamical systems in the variational formulation of the Fokker–Planck equation by the Wasserstein metric*, Appl. Math. Optim. **42** (2000), 203–227.
- [69] R.H. Nochetto, G. Savaré and C. Verdi, *A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations*, Comm. Pure Appl. Math. **53** (2000), 525–589.
- [70] F. Otto, *Doubly degenerate diffusion equations as steepest descent*, Preprint, Univ. of Bonn (1996).
- [71] F. Otto, *Dynamics of labyrinthine pattern formation in magnetic fluids: A mean-field theory*, Arch. Ration. Mech. Anal. **141** (1998), 63–103.
- [72] F. Otto, *Lubrication approximation with prescribed nonzero contact angle*, Comm. Partial Differential Equations **23** (1998), 2077–2164.
- [73] F. Otto, *Evolution of microstructure in unstable porous media flow: A relaxational approach*, Comm. Pure Appl. Math. **52** (1999), 873–915.
- [74] F. Otto, *The geometry of dissipative evolution equations: The porous medium equation*, Comm. Partial Differential Equations **26** (2001), 101–174.
- [75] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361–400.
- [76] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal. **37** (2005), 1227–1255.
- [77] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer-Verlag, New York (1983).
- [78] M. Pierre, *Uniqueness of the solutions of $u_t - \Delta\varphi(u) = 0$ with initial datum a measure*, Nonlinear Anal. **6** (1982), 175–187.
- [79] A. Pratelli, *On the equality between Monge’s infimum and Kantorovich’s minimum in optimal mass transportation*, Ann. Inst. H. Poincaré Probab. Statist., to appear.
- [80] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin (1998).
- [81] R. Rossi and G. Savaré, *Gradient flows of non convex functionals in Hilbert spaces and applications*, ESAIM Control Optim. Calc. Var. **12** (2006), 564–614.
- [82] J. Rulla, *Error analysis for implicit approximations to solutions to Cauchy problems*, SIAM J. Numer. Anal. **33** (1996), 68–87.
- [83] C. Sparber, J.A. Carrillo, J. Dolbeault and P.A. Markowich, *On the long-time behavior of the quantum Fokker–Planck equation*, Monatsh. Math. **141** (2004), 237–257.
- [84] K. Sturm, *On the geometry of metric measure spaces*, Acta Math., to appear.
- [85] C. Villani, *Optimal transportation, dissipative PDE’s and functional inequalities*, Optimal Transportation and Applications, Martina Franca (2001); Lecture Notes in Math., Vol. 1813, Springer-Verlag, Berlin (2003), 53–89.
- [86] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. Math., Vol. 58, Amer. Math. Soc., Providence, RI (2003).

CHAPTER 2

The Mathematics of Chemotaxis

M.A. Herrero

*Departamento de Matemática Aplicada, Facultad de CC. Matemáticas,
Universidad Complutense de Madrid, Avda. Complutense s/n, 28040 Madrid, Spain
E-mail: Miguel_Herrero@mat.ucm.es*

Contents

1. Introduction: What is chemotaxis?	139
2. How do chemotactic units work?	141
2.1. Ligand binding to receptors	143
2.2. The role of fluctuations	144
2.3. Diffusion effects on ligand binding	147
2.4. Estimating the measurement error	149
2.5. Receptor clustering	154
2.6. Signaling pathways and their performance properties	157
3. Some mathematical problems arising from the study of <i>Dictyostelium discoideum</i>	168
3.1. The social life of <i>Dictyostelium discoideum</i> in a nutshell	169
3.2. Early aggregation stages: the Keller–Segel model	170
3.3. The Keller–Segel model revisited: from micro to macro	175
3.4. Pattern formation in <i>Dictyostelium discoideum</i>	180
Acknowledgements	189
References	189

Abstract

This chapter provides a description of some of the mathematical approaches that have been developed to account for quantitative and qualitative aspects of chemotaxis. This last is an important biological property, consisting in motion of cells induced by chemical substances, which is known to occur in a large number of situations, both homeostatic and pathological. Particular attention will be paid to the limits on a cell's capability to measure external cues on the one hand, and to provide an overall description of aggregation models for the slime mold *Dictyostelium discoideum* on the other.

1. Introduction: What is chemotaxis?

Chemotaxis is a technical term which is commonly used to describe the motion of cells induced by chemical substances, either to navigate toward the source of those, or else to escape away from them. This scientific concept was formulated as a hypothesis by Ramón y Cajal [94] in 1893, in the course of his seminal studies on Developmental Neuroscience. As recalled for instance in [105] and [27], Cajal noticed the remarkable behavior of growing axons of neurons, which maintain a precise orientation toward their target cells during their growth. As a matter of fact, for a mature nervous system to work properly, a very precise pattern of connections among a huge number of neurons (about 10^{12} in human brain) has to be established during embryonic and early postnatal periods, and such connections are made by migration of neurons from their proliferative sites to their eventual targets. Cajal considered what the mechanism could be for the “intelligent force” responsible for such guiding process. He eventually formulated in [94] the so-called neurotropic (or quimiotactic) theory, according to which target cells secrete attracting substances, and growing neuronal axons possess a chemotactic sensitivity (a “chemically induced ameboidism”) that allow them to follow their way in the course of their motion toward their final destination (see Figure 1). In his later work on regeneration of the nervous system [96], Cajal observed that, in sectioned nerves, regenerating peripheral axons arising from proximal stumps will always go toward distal stumps, even if considerable obstacles are raised against their growth.

While these facts provided considerable support to the assumption of chemotactic guidance in neural navigation, the identification of the first molecules with chemotropic action in mammalian embryos took place a century afterward (cf. [104]). As a matter of fact, in a series of studies performed after 1980, a number of key features of neural wiring were ascertained, as for instance, the existence of intermediate targets that assist in keeping ax-

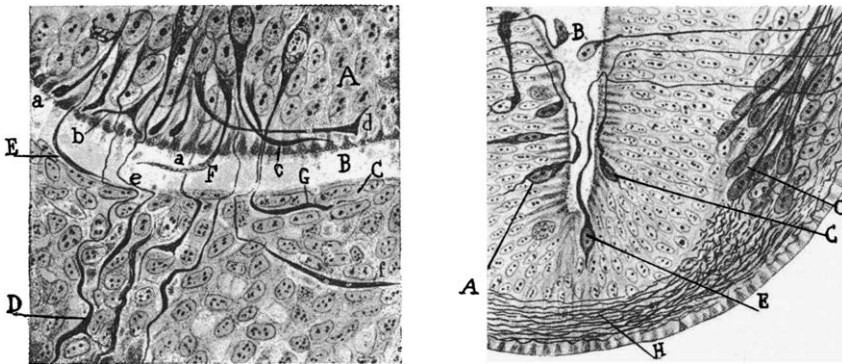


Fig. 1. Left: A section of early bone marrow (A) and mesodermic tissue taken from a three-day duck embryo. Notice that in younger neuroblasts growth cones always proceed in between the cells; E, F – growth cones freely moving through perimedullary space; D – growth cones already placed at the mesodermal area. Right: Hystological preparation corresponding to the bulb of a four-days chicken embryo. Note that nervous fibers accidentally placed at the ventricule (A, E, C) appear to be free, and their axons orient themselves to move toward their destination through the nervous field. (Reproduced from [95].)

ons en route toward their final destinations [113], and the existence of repulsive molecules, that keep moving cells away from unwanted places [20]. In this way a scenario emerged in which neural navigation proceeds according to short-range cues and long-range, diffusive signals that may be either attractive or repulsive.

While neuroscience is arguably the field in which chemotaxis was first postulated, during the XX century a wealth of evidence has been accumulated, and considerable insight has been gained, on the role played by this type of directed motion in the life of a number of species.

One of these is the slime mold *Dictyostelium discoideum* (*Dd* for short), a kind of amoebae first identified by Raper in 1935 (cf. [97] and also [15]). These are nucleated cells that live in forest soils, feeding on bacteria. As long as there is sufficient food supply, *Dd* cells have a life cycle analogous to other microorganisms sharing the same habitat, and they proliferate by cell fission. However, when food becomes scarce, the social life of *Dd* amoebae is dramatically altered. To begin with, some cells start emitting pulses of a chemical (adenosine 3',5'-cyclic monophosphate, cAMP) which acts as a communication signal. Cells are able to receive it, transduce it and then, after internally producing cAMP, they release it outside, thus keeping a cAMP feedback loop. As a consequence of this signaling process, amoebae eventually aggregate into mounds of rather constant size (up to 10^5 units), to form multicellular pseudoorganisms. These last subsequently enter into a developmental program which involves cell differentiation and migration, to eventually form a fruiting body which consists in a ball of spores (resistance forms that may remain in a quiescent state for several weeks) located on top of a thin stalk (for recent surveys, see [34,48,73], etc.). As observed for instance in [82], this clear separation in time from aggregation and differentiation makes *Dd* a suitable model organism to study in vivo both processes, that quite often occur simultaneously in other species.

Another biological model on which chemotaxis has been extensively studied is *Escherichia coli* (*E-coli*), a bacteria that usually colonizes the human bowel a few hours after birth, and which may be responsible for a number of serious infectious diseases. *E-coli* is able to swim toward sources of chemoattractants as aspartate or glucose by using as propellers the flagella it is provided with; see, for instance, [10] for a comprehensive description of *E-coli* behavior. White blood cells (and in particular neutrophils) provide a further example of chemotactic cells. For instance, neutrophils are known to navigate comparatively long distances to arrive at places in the body where injuries occur. An account of the physiological mechanisms which mediate chemotaxis in these (and others) types of cells can be found in [31], a monograph we refer to for additional information.

In spite of the considerable differences among such types of organisms (for instance, diameter lengths in *Dd* and *E-coli* differ by one order of magnitude), chemotactic cells present some common rules of functioning, governed by physical processes that may change from one species to other. To begin with, they should be endowed with a fine sensitivity, thus being able to detect small changes in chemoattractant concentrations around them. At the same type they should show adaptation, and therefore remain largely indifferent to important changes in homogeneous concentrations of otherwise stimulant chemicals. Once a (gradient) signal has been detected, they should be able to build up an amplified, internal signal transmission network strong enough to reorient its movement and yet sufficiently flexible to change direction again when necessary. Finally, in many instances

chemotactic migration involves coordinated motion of a large number of cells, which is regulated by intense intercellular communication.

In this chapter we shall review some of the quantitative models that have been proposed to account for some of the aforementioned aspects of chemotaxis. The term quantitative refers both to the physical assumptions, and the mathematical formulation thereof, which are advanced as a tool to gain insight in the way chemotaxis proceeds. As it will be apparent from our forthcoming discussion, quantitative modeling in chemotaxis is currently at a preliminary, although promising, level. Out of the various relevant features in chemotaxis, only a few will be addressed here, according to the plan which is briefly described below.

In Section 2 we shall be concerned with the actual manner in which individual chemotactic cells operate, and the limitations (and success stories) of such procedures will be described. In short, the line of thought goes as follows: cells detect changes in chemical gradients around them by monitoring the state of occupancy of specific receptors at their membranes, and they make use of the information thus gathered to trigger internal signaling cascades. Among other things, these last eventually result in cells oriented motion toward their targets. The discussion made in Section 2 deals therefore with individual cell behavior and their operating limits, and in this sense is not restricted to chemotactic cells, although the approaches to be described below have largely arisen in a chemotaxis setting.

Going from general to particular, in Section 3 we examine some of the mathematical problems whose study has been motivated by particular aspects of the cell cycle of *Dictyostelium discoideum*. As will be remarked then, some of these problems have become topics of mathematical interest in their own, so that their relation to the biological source has become fainter. In any case an attempt has been made to keep in mind the biological motivation as often as possible. A major issue to be addressed therein is pattern formation in *Dd* cultures. This is merely a part, however interesting, of a vast subject in which a huge literature is available. We have chosen to focus on a few topics, where mathematics has played a relevant role (and has benefited much from consideration of the problems involved). In few words, the questions examined in that section deal with early aggregation properties of *Dd* and the target-spiral transition which is customarily seen to mediate the establishment of aggregation centers.

In selecting such a reduced number of issues to address, many interesting features are left out; a particularly interesting example of this omission is cell motility. A second one (in a rather long list) concerns three-dimensional aspects of *Dd* culmination, of which only a few words will be said below. To such limitations in choice, restrictions in style will be added. Indeed, our approach will be basically descriptive, without going into the detail of mathematical proofs or arguments; for this we refer to the original sources where appropriate.

2. How do chemotactic units work?

To move toward a distant location, a cell should first receive a chemical cue released therein (or at some intermediate destination). This signal has then to be processed, to derive information about the position of the target it will eventually travel to. The first step in this process involves the interaction of signaling molecules (ligands) with specific receptors

located on the cell surface. These consist in macromolecules transversally inserted at the cell membrane. They thus possess an extracellular domain, where ligands land in, and an intracellular one. This last is instrumental in transducing signals by means of chemical processes (phosphorylation, methylation, ...) to trigger the subsequent cellular response.

It is natural to assume that a cell obtains information about, say, a source of chemoattractant from the state of occupancy of its membrane receptors (we say that a receptor is occupied as long as a ligand remains bound to it). To proceed in an efficient way, a cell should be able to monitor from that raw data the concentration of chemoattractant in its neighborhood. In particular, it has to detect small variations in its gradients (which may change their space distribution as time passes), a property often termed as sensitivity. Moreover, a fine sensitivity should go hand in hand with a high signal amplification downstream in the chemical cascade started by ligand binding to receptors. Amplification is required to set in motion the inner machinery of the cell, which will eventually result in navigation to a chemical target. Since cells move in media where chemical concentrations may vary over several orders of magnitude, it is necessary that the detection process be independent of the absolute concentration of isotropic ligands, a fact usually referred to as adaptation.

What does a cell know about the world around it? The physical limits to what a cell can actually measure were examined in a seminal article by Berg and Purcell in 1977 [11]. Some of the key points addressed in that work were succinctly described by the authors at their Introduction:

“... In the world of a cell as small as a bacterium, transport of molecules is effected by diffusion, rather than bulk flow; movement is resisted by viscosity, not inertia; the energy of thermal fluctuation, kT , is large enough to perturb the cell's motion. In these circumstances, what are the physical limitations on the cell's ability to sense and respond to changes in its environment? What, for example, is the smallest change in concentration of a chemical attractant that a bacterium could be expected to measure reliably in a given time?”

In this section we shall review some of the mathematical approaches that have been proposed to gain insight into the key problem raised above: the manner in which a cell obtains information from receptor occupancy by ligand binding. To this end, we proceed in several steps. To start with, we shortly recall in the next subsection a basic model for ligand binding according to mass action law. This is done in an isotropic setting, and a formula for the equilibrium concentration is provided. We then discuss in Section 2.2 the role played by fluctuations in the distribution of receptor occupancy. In particular, it will be observed that, while ligand fluctuations are likely to be negligible, those in the kinetic binding process are rather important instead.

In Section 2.3 we consider the effect of diffusion (a particularly relevant type of mass transport) in the influx of ligands to a cell's surface. Of particular interest is the formula therein provided for the total current of ligands that may be absorbed by a system of N circular receptor patches scattered over the surface of a spherical cell (cf. (37), (38)). We then discuss in Section 2.4 the question of error measurement in two possible methods to measure ligand concentration from receptor occupancy: these are respectively known as spatial gradient and temporal gradient sensing. As a result, crude (but illuminating) estimates will be provided for the minimum gradient that can be detected, and the longest distance that a cell can navigate.

Impressive as these predictions are, cells are known to do even better. This they do by resorting to mechanisms only partially explored as yet. One of these is receptor clustering, which shall be dealt with in Section 2.5, where interaction of bivalent ligands with cell receptors is described in some detail, and analogies with polymerization processes are noticed. Understanding the precise manner in which clustering results in higher efficiency seems to require, however, of a detailed knowledge of the functioning of intracellular signaling pathways. This important topic will be dealt with in Section 2.6.

2.1. Ligand binding to receptors

In this and forthcoming subsections we shall borrow from a basic monograph due to Lauffenburger and Linderman [65], where fundamental aspects of receptor operation are described in detail. In its simplest setting, the process under consideration involves a monovalent ligand L that reversibly binds to a monovalent receptor R , to form a receptor–ligand complex C ,



where k_f (respectively k_r) denotes the kinetic rate of binding (respectively dissociation) of the process under consideration. According to the mass action law, a mathematical model for (1) is given by

$$\frac{dC}{dt} = k_f RL - k_r C. \quad (2)$$

To (2), some conservation laws should be added. In particular, the total number of receptors has to be preserved,

$$R + C = R_T, \quad (3)$$

and in many cases the amount of ligand may be assumed to remain unchanged during the process. If ligand concentration is measured in moles per volume, receptors are measured in number per cell and cells are present at concentration n (number per volume), one then has that

$$L + \frac{n}{N_A} C = L_0, \quad (4)$$

where $N_A = 6.02 \times 10^{23}$ is Avogadro's constant, giving the number of molecules per mole. If we further assume that $(n/N_A)C \ll L_0$, equations (2)–(4) simplify to

$$\frac{dC}{dt} = k_f(R_T - C)L_0 - k_r C, \quad (5)$$

that can be integrated to yield

$$C(t) = C_0 \exp(-(k_f L_0 + k_r)t) + \frac{k_f L_0 R_t}{k_f L_0 + k_r} (1 - \exp(-(k_f L_0 + k_r)t)). \quad (6)$$

Equation (6) describes rapid equilibration toward the steady-state value

$$C_{eq} = \frac{R_T L_0}{K_D + L_0} \quad \text{with } K_D = \frac{k_r}{k_f}, \quad (7)$$

where K_D is termed the equilibrium dissociation constant.

2.2. The role of fluctuations

In the previous deterministic model, the level of receptor occupancy is described by the formation of complexes C . However, a number of random factors may alter the values thus obtained. For example, random fluctuations in the ligand concentration near a cell may result in deviations from the values predicted by formulae (6) and (7). Following [65], we consider the effect of ligand random fluctuations on the equilibrium formula (7). The corresponding fluctuation in the number of complexes formed is given by

$$\delta C_{eq} = \frac{dC_{eq}}{dL} \delta L = \frac{R_T K_D}{(K_D + L)^2} \delta L, \quad (8)$$

where δC_{eq} denotes the standard deviation in C_{eq} as a result of a standard deviation δL in the value of L . The relative magnitude of these fluctuations in receptor binding is

$$\frac{\delta C_{eq}}{C_{eq}} = \left(1 + \frac{L}{K_D}\right)^{-1} \frac{\delta L}{L}. \quad (9)$$

To estimate $\delta C_{eq}/C_{eq}$ one thus needs an estimate for $\delta L/L$, and this last can be obtained from the formula

$$\frac{\delta L}{L} = (N_A L V)^{-1/2} \quad (10)$$

(cf. [11] and [65]). Here V is the volume of the medium accessible for ligand binding, and $N_A L V$ is the expected number of ligand molecules in that volume. A natural choice is $V \sim l^3$, where l is a characteristic length of the medium. For instance, if ligand transport is assumed to occur by a diffusion process with diffusivity D_L , then the distance traveled by an average molecule in a time t_* will be $l \sim (D_L t_*)^{1/2}$. If we take $t_* = k_r^{-1}$ (that is, the mean time period between receptor binding events), then at $L = K_D$ we would obtain

$$\frac{\delta L}{L} \sim (N_A (D_L k_r^{-1})^{3/2} K_D)^{-1/2}. \quad (11)$$

As pointed out in [65], for values $D_L \sim 10^{-6}$ – 10^{-5} cm²/s, $k_r = 10^{-4}$ – 10^{-1} s⁻¹, $K_D = 10^{-10}$ – 10^{-6} moles/volume, (11) yields $\delta L/L \sim 10^{-7}$ – 10^{-2} , which by (9) translate into an estimate for $\delta C_{eq}/C \sim 10^{-5}$ to 1%.

Consider next the case of fluctuations in the kinetic binding process. In this context, rate constants may be given a probabilistic meaning. For instance, k_r can be thought of as the probability that a single complex will dissociate. Therefore, for $0 < \delta t \ll 1$, the probability of a single dissociation event at a given receptor will be $k_r \delta t$. More precisely, let $P_j(t)$ be the probability that there are j complexes on a cell at a time t . The change in the number of complexes occurring in a time interval δt , $0 < \delta t \ll 1$, assuming that there were C complexes at time t , is described by the kinetic equation

$$\begin{aligned} P_C(t + \delta t) - P_C(t) &= k_f L (R_T - (C - 1)) P_{C-1}(t) \delta t \\ &\quad - k_f L (R_T - C) P_C(t) \delta t \\ &\quad - k_r C P_C(t) \delta t + k_r (C + 1) P_{C+1}(t) \delta t. \end{aligned} \quad (12)$$

In the limit $\delta t \rightarrow 0$, this leads to

$$\begin{aligned} \frac{dP_C(t)}{dt} &= k_f L (R_T - (C - 1)) P_{C-1} + k_r (C + 1) P_{C+1} \\ &\quad - (k_f L (R_T - C) + k_r C) P_C(t) \end{aligned} \quad (13)$$

for $C = 1, 2, \dots, R_T - 1$. This set of equations is to be completed with

$$\frac{dP_0}{dt} = -k_f L R_T P_0 + k_r P_1, \quad (14)$$

$$\frac{dP_{R_T}}{dt} = k_f L P_{R_T-1} - k_r R_T P_{R_T}. \quad (15)$$

Following a standard terminology, the set of equations (13)–(15) is customarily termed as the master equation for the process under consideration. To solve it, a classical method consists in introducing a generating function [38]

$$G(s, t) = \sum_{C=0}^{R_T} s^C P_C(t), \quad (16)$$

so that

$$P_0(t) = G(0, t), \quad P_C(t) = \frac{1}{C!} \left[\frac{d^C G}{ds^C} \right]_{s=0} \quad \text{for } C = 1, \dots, R_T - 1. \quad (17)$$

As a matter of fact, on multiplying each equation in (13)–(15) by s^C and then adding them up, one eventually obtains

$$\frac{\partial G}{\partial t} = (1 - s) \left((k_f L s + k_r) \frac{\partial G}{\partial s} - k_f L R_T G \right). \quad (18)$$

To (18), the initial condition $P_C(0) = 0$ for $C \neq 0$, $P_C(0) = 1$ for $C = 0$ has to be added, which in terms of G reads

$$G(s, 0) = 1. \quad (19)$$

Furthermore, the requirement that the sum of all probabilities be equal to one yields

$$G(1, t) = 1. \quad (20)$$

In many instances, we are interested in the mean value of C , denoted $\langle C \rangle$, and the variance σ_C^2 . A quick check reveals that

$$\langle C \rangle = \sum_{C=0}^{R_T} C P_C = \left[\frac{\partial G}{\partial s} \right]_{s=1}, \quad (21)$$

$$\sigma_C^2 = \sum_{C=0}^{R_T} (C - \langle C \rangle)^2 P_C = \left[\frac{\partial^2 G}{\partial s^2} + \frac{\partial G}{\partial s} - \left(\frac{\partial G}{\partial s} \right)^2 \right]_{s=1}. \quad (22)$$

Consider now the case of the steady-state solution of (18), obtained by setting $\partial G / \partial t = 0$ therein. We then may solve the resulting equation by direct integration, and then use (21), (22) to obtain

$$\langle C_{\text{eq}} \rangle = \frac{R_T L}{K_D + L}, \quad (23)$$

$$\delta C_{\text{eq}} \equiv (\sigma_C)_{\text{eq}} = \frac{(R_T L K_D)^{1/2}}{K_D + L}, \quad (24)$$

the last estimate being proportional to the total number of cell receptors R_T . From (23) and (24) it follows that

$$\frac{\delta C_{\text{eq}}}{C_{\text{eq}}} = \left(\frac{K_D}{L R_T} \right)^{1/2}.$$

In particular, when $L = K_D$ we obtain

$$\frac{\delta C_{\text{eq}}}{C_{\text{eq}}} = R_T^{-1/2}. \quad (25)$$

For instance, for $R_T = 10^4$ receptors/cell, statistical fluctuations with relative magnitude of 1% are expected, a value which falls well within the sensitivity threshold known for chemotactic cells [5,121]. On the other hand, a comparison of (25) with the corresponding value obtained at the end of our previous subsection reveals that this type of fluctuations is more likely to have an impact on chemotaxis than the previous one.

Concerning the time-dependent equation (18), we point out that the corresponding solution that satisfies (19) and (20) can be readily obtained by integration along characteristics (cf., for instance, [55]). Using then (21) and (22), it follows that

$$\langle C(t) \rangle = \frac{R_T L}{K_D + L} (1 - \exp(-(k_f L + k_r)t)), \quad (26)$$

$$\begin{aligned} (\sigma_C^2)_{\text{eq}} &= \frac{R_T L}{(K_D + L)^2} (L \exp(-(k_f L + k_r)t) + K_D) \\ &\quad \times (1 - \exp(-(k_f L + k_r)t)). \end{aligned} \quad (27)$$

2.3. Diffusion effects on ligand binding

We have already noticed that diffusion has a limited influence on ligand fluctuations near the cells. However, this type of mass transfer process is the dominant mechanism to carry ligands toward the cells surface, so that reactions as that described in (1), (2) may occur. We shall briefly recall below some quantitative aspects of the role played by diffusion in the arrival of chemical signals at a cell.

To begin with, let us consider the following auxiliary problem. To determine the steady-state concentration of a ligand away from a single spherical cell, which is centered at the origin ($r = 0$) and whose surface corresponds to $r = R > 0$. One is thus led to solve

$$D \frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dL}{dr} \right) = 0 \quad \text{for } r > R. \quad (28)$$

Assuming constant concentration away from the cell, we require

$$L \rightarrow L_0 > 0 \quad \text{as } r \rightarrow \infty. \quad (29)$$

To solve (28), (29) an additional boundary condition has to be imposed at $r = R$. For later reference, we state below a choice of particular interest:

$$I \equiv 4\pi R^2 D \left[\frac{dL}{dr} \right]_{r=R} = k_{\text{on}} L(r) \quad \text{at } r = R, \quad k_{\text{on}} \geq 0. \quad (30)$$

It is easy to check that the solution to (28)–(30) is provided by

$$L(r) = -\frac{k_{\text{on}} R L_0}{4\pi D R + k_{\text{on}}} \frac{1}{r} + L_0. \quad (31)$$

Suppose now that the whole surface of the cell is covered by receptors, and that ligands are instantly absorbed as soon as they arrive there. In this case, (30) has to be replaced by

$$L(r) = 0 \quad \text{at } r = R, \quad (32)$$

and the solution to (28), (29) and (32) is now given by

$$L(r) = L_0 \left(1 - \frac{R}{r} \right) \quad \text{for } r > R, \quad (33)$$

so that the total flux carried out to $r = R$ is

$$I_M \equiv 4\pi R^2 D \left[\frac{dL}{dr} \right]_{r=R} = 4\pi R D L_0 \equiv k_+ L_0. \quad (34)$$

Relation (34) defines the observable association rate constant k_f . A second limit case of interest corresponds to the situation where there is only a single, circular receptor with radius s , $0 < s \ll R$, located over the cell surface ($r = R$), which is impervious to ligands except for the receptor patch. In this case (30) has to be replaced by

$$L = 0 \quad \text{at the receptor}, \quad \frac{\partial L}{\partial r} = 0 \quad \text{otherwise over } r = R. \quad (35)$$

The solution to (28), (29) and (35), while still explicit, is not nearly as straightforward as that given by (31) or (33) (cf. [11] and [25], p. 42). The corresponding flux is now given by

$$4\pi R^2 D \left[\frac{\partial L}{\partial r} \right]_{r=R} = 4DsL_0. \quad (36)$$

Bearing in mind the two extreme cases (32) and (36), an asymptotic formula was derived in [11] corresponding to the case where the number of receptors N is large ($N \gg 1$), but the average distance d_s among them satisfies $d_s \gg s$, so that they are fairly separated from each other. A geometrical argument reveals that this is the case if $N^{1/2}s \ll R$. The estimate obtained in [11] reads

$$I = I_M \frac{Ns}{Ns + \pi R}. \quad (37)$$

Actually, a correction to formula (37) was later provided in [122] by means of an effective-medium argument, namely

$$I = I_M \frac{Ns}{Ns + \pi R(1 - p_A)}, \quad p_A = \frac{N\pi s^2}{4\pi R^2}, \quad (38)$$

so that p_A represents the fraction of the sphere's surface which is covered with circular receptors. A striking consequence of (37) is that a large incoming flux can be achieved with relatively few, well-separated receptors. For instance, according to (37), $I = I_M/2$ if $N = \pi R/s$. If $R = 5 \mu\text{m}$ ($1 \mu\text{m} = 10^{-6}$ meters) and $s = 10 \text{ \AA}$ ($1 \text{ \AA} = 10^{-10}$ meters), the value $I_M/2$ is achieved when $N \sim 15,700$, the average distance among receptors is $1400 \text{ \AA}$, and only a fraction of about 10^{-4} of the cell surface is covered by receptors [9].

Let us elaborate a bit on some of the formulae previously obtained. To this end, we follow [65] and observe that the binding of two molecules, as denoted by (1), is, in fact, a two-step process. First, molecular transport of the species R and L is required; we denote the corresponding rate constant by k_f . In our case, that transport is assumed to be due to diffusion. Then a chemical reaction takes place, which is characterized by the intrinsic association (respectively, dissociation) rate k_{on} (respectively, k_{off}). Thus the kinetic constants k_f, k_r in (1) are actually combinations of k_f, k_{on} and k_{off} described before. Moreover, constant D in (28) is such that $D = D_L + D_R$, the sum of ligand and receptor diffusivities. On the other hand, if a receptor is present at $r = R$, one has to write $k = k_{on}$ in (30). In this case, the forward rate k_f is given by

$$k_f = L_0^{-1} 4\pi R^2 D \left[\frac{dL}{dr} \right]_{r=R}. \quad (39)$$

Recalling the definition of k_{on}, k_+ in (30) and (34) respectively, one deduces from the previous remarks and (30) that

$$k_f = \frac{4\pi DR k_{on}}{4\pi DR + k_{on}} \equiv \frac{k_+ k_{on}}{k_+ + k_{on}} = \left(\frac{1}{k_+} + \frac{1}{k_{on}} \right)^{-1}. \quad (40)$$

As pointed out in [65], this formula allows for an appealing interpretation: the overall resistance to binding, denoted by $1/k_f$, is the sum of the resistance to diffusion $1/k_+$ and that to reaction $1/k_{on}$. In particular, if $k_{on} \gg k_+$, $k_f \sim k_+ = 4\pi DR$ and the binding is termed diffusion-limited. Conversely, when $k_{on} \ll k_+$, $k_f \sim k_{on}$ and the binding is considered to be reaction-limited.

2.4. Estimating the measurement error

In general, chemotactic cells move along paths for which their receptor occupancy gradient (spatial or temporal) is maximum. However, changes in occupancy are often so small that they hardly can be distinguished from fluctuations inherent to ligand binding. One many therefore wonder what are the physical limits imposed on a cell's ability to detect a chemical gradient.

To address this issue, we shall take up the analysis introduced in [11] and then developed in [28,29]. We shall roughly proceed as follows. One first assumes the incoming signal to be a function of the receptor occupancy. Then an estimate on the standard deviation about the mean signal is obtained, which is in turn used to derive a lower bound on gradient detection.

Consider first the case of a hypothetical mechanism based on spatial gradient detection. Such procedure requires estimating occupancy variations along a dimension parallel to the gradient. As before, we denote by L the ligand concentration around a cell, and write $p(L, t)$ to represent the associated fractional receptor occupancy. It is natural to assume that p arises as an average of a random variable describing ligand binding to receptors. If the cell is assumed to be spherical, and the concentration change across its diameter is given

by ΔL , the changes in ligand concentration and occupancy across a diameter will respectively be given by $(L + \Delta L)$ and $p(L + \Delta L, t)$. For a (nearly) constant gradient $\partial L/\partial x$, we have that

$$\Delta L \sim R \frac{\partial L}{\partial x}, \quad (41)$$

R being the cell's diameter. As a measure of the signal received, we may take

$$S = \frac{1}{T} \int_0^T (p(L + \Delta L, t) - p(L, t)) dt, \quad (42)$$

where $T > 0$ is an averaging time. The minimal requirement for a gradient to be detected is

$$S > \sigma, \quad (43)$$

where $\sigma(L, t)$ is the standard deviation in the measured occupancy, which is given by

$$\sigma^2(L, t) = \left\langle \left(\frac{1}{T} \int_0^T p(L, t) dt \right)^2 \right\rangle - \left\langle \frac{1}{T} \int_0^T p(L, t) dt \right\rangle^2. \quad (44)$$

Estimating σ^2 in (44) involves dealing with the corresponding autocorrelation functions [11,28]. Arguing as in [28], one may show that

$$\sigma^2(L, t) = \frac{2}{NT^2} \int_0^T dt \int_0^t p(s)(1 - p(s)) \exp\left(-\int_s^t \frac{dn}{\tau}\right) ds,$$

where N is the number of receptors per cell, $p(s)$ is the fractional occupancy at time s and τ is the relaxation time for ligand–receptor binding (which also depends on time). From (41), and assuming $\Delta L \ll L$, one readily sees that

$$p(L + \Delta L) - p(L) \sim R \frac{\partial L}{\partial x} \frac{\partial p}{\partial L}, \quad (45)$$

whence

$$S \sim \frac{R}{T} \frac{\partial L}{\partial x} \int_0^T \frac{\partial p}{\partial L} dt,$$

and condition (43) reads

$$\left(\frac{R}{T} \frac{\partial L}{\partial x} \int_0^T \frac{\partial p}{\partial L} dt \right)^2 > 2\sigma^2(L, t) \quad (46)$$

(cf. [29]), where we have made use of the assumption $\sigma^2(L, t) \sim \sigma^2(L + \Delta L, t)$ for $\Delta L \ll L$.

A second detection mechanism is based in measuring temporal gradients. These arise as cells move through a spatial gradient. In this case, in order to measure the signal we consider the expression

$$S = \frac{1}{T} \int_{t_1}^{T+t_1} p(L + \Delta L, t) dt - \frac{1}{T} \int_0^T p(L, t) dt, \quad (47)$$

where $\Delta L \sim vt_1 \partial L / \partial x$ and v denotes the cell velocity through the spatial gradient $\partial L / \partial x$. From (47) and (43) we thus obtain a condition for temporal gradient detection,

$$\begin{aligned} & \frac{1}{T^2} \left(\int_{t_1}^{T+t_1} p(L, t) dt + vt_1 \frac{\partial L}{\partial x} \int_{t_1}^{T+t_1} p(L, t) dt - \int_0^T p(L, t) dt \right)^2 \\ & > \sigma^2(L, T + t_1) - \sigma(L, t). \end{aligned} \quad (48)$$

Consider now the case of chemical equilibrium. Then there holds

$$p = \frac{KL}{1 + KL}, \quad \text{where } K = K_D^{-1} = \frac{k_f}{k_r} \quad (49)$$

(see (7)), so that

$$\frac{\partial p}{\partial L} = \frac{K}{(1 + KL)^2}, \quad (50)$$

and (44) yields now

$$\sigma^2(L, t) = \frac{2KL\tau}{NT(1 + KL)^2}, \quad \frac{1}{\tau} = k_f L + k_r \quad (51)$$

(cf. (6) for the second statement above). Since $\tau/T \rightarrow 0$ at equilibrium, condition (46) gives

$$\frac{T}{\tau} > \frac{4(1 + KL)^2}{KLN} \left(\frac{R}{L} \frac{\partial L}{\partial x} \right)^{-2} \equiv u_s. \quad (52)$$

Note that u_s in (52) can be thought of as the minimum value of T/τ needed to detect a spatial gradient, under the assumption that equilibrium is rapidly arrived at.

For temporal gradient detection instead, and assuming that (49) holds, the first and third integrals in the left-hand side of (48) cancel out, and we obtain

$$S^2 = \left(vt_1 \frac{\partial L}{\partial x} \frac{\partial p}{\partial L} \right)^2 = \left(\frac{kv t_1}{(1 + KL)^2} \frac{\partial L}{\partial x} \right)^2 = \left(\frac{K t_1}{(1 + KL)^2} \frac{\partial L}{\partial t} \right)^2. \quad (53)$$

Taking into account (51) and (53), we obtain the condition

$$\frac{T}{\tau} > \frac{4(1 + KL)^2}{KLN} \left(\frac{t_1}{L} \frac{\partial L}{\partial t} \right)^{-2} \equiv \left(\frac{R}{vt_1} \right) u_s, \quad (54)$$

where u_s is as in (52).

We are now ready to compare the relative efficiency of both detection mechanisms previously described. Consider first the case of spatial gradient detection, and assume that at chemical equilibrium n more receptors are occupied on the high concentration side of the cell than on the low concentration one. Then, from (45) and (50) it follows that

$$n = R \frac{\partial L}{\partial x} \frac{KN}{(1 + KL)^2}.$$

From this and (52) one has

$$\bar{u}_s = \frac{4N}{n^2} \frac{KL}{(1 + KL)^2}. \quad (55)$$

Since $f(x) = x(1 + x)^{-2}$ has a maximum at $x = 1$, we readily see from (55) that $\bar{u}_s$ has a maximum at $KL = 1$. When the number of receptors N is of order $N \sim 10^4 - 10^5$, we thus obtain:

$$\frac{10^4}{n^2} \leq u_s \leq \frac{10^5}{n^2}. \quad (56)$$

As observed in [29], (56) sets a severe limitation on the possibility of detecting small occupancy differences. Suppose for instance that $n \sim 10$. Then $u_s \sim 10^2 - 10^3$, and according to (52) the averaging time T will be of the order of seconds if $\tau \sim 10^{-3}$ s, and in the range of hours to days if τ is of the order of seconds to minutes. When we particularize to bacterial cells as *E-coli*, the first situation is known to occur (that is, $\tau \sim 10^{-3}$ s) which in view of (51) requires dissociation constants of the order of 10^{-3} s $^{-1}$. Since a typical reaction-limited forward rate lies in the order of $10^5 - 10^6$ s $^{-1}$, only low-affinity ligand binding would be allowed in this case.

Let us examine now the case of temporal gradient detection. Assuming again chemical equilibrium, one obtains from (49) and (47) that

$$S \sim \frac{K \nu t_1}{(1 + KL)^2} \frac{\partial L}{\partial x}.$$

Recalling (51) and (48), signal and noise are now comparable when

$$T \sim 4\tau \left(\frac{R}{\nu t_1} \right)^2 \frac{KLN}{(1 + KL)^2 n^2}. \quad (57)$$

Consider for instance the case where $KL = 1$, $N \sim 3 \times 10^3$. Then (57) yields $T \sim 3 \times 10^4 \tau (R/\nu t_1 n)^2$, which is to be compared with $T \sim 3 \times 10^4 \tau / n^2$ obtained in the case of spatial gradient sensing (see (54)). As pointed out in [29], this example shows that for $\nu t_1 \sim R$, temporal sensing offers no advantage over spatial sensing. A further point

to be noticed is that for a temporal sensing mechanism, the time T_0 required to detect a gradient is

$$T_0 \sim T + t_1,$$

which in view of (57) reads

$$T_0 \sim \frac{4\tau K L N R}{(1 + K L)^2 v} \frac{1}{n^2 t_1^2} + t_1. \quad (58)$$

Since, for $B > 0$, $g(t_1) = B/t_1^2 + t_1$ achieves a minimum at $t_1 = (2B)^{1/3}$, (58) provides a way of estimating the minimum time to detect a gradient by a temporal mechanism. To that end, the key parameter turns out to be the ratio $(NR^2\tau/v^2n^2)$. As discussed in [29], the picture that emerges can be roughly described as follows. For small bacterial cells, the temporal mechanism permits detection of affinities 10^2 – 10^3 higher than could be obtained from a spatial mechanism (see Figure 3 in [29]). However, for large crawling cells the affinity range on which temporal detection fares better is much more restricted, and spatial detection becomes more efficient for low affinity ligands (Figure 4 in [29]).

What is the maximum distance that a cell can navigate in the trail of a chemical scent? A simple estimate can be provided by requiring that the ligand should have a relative concentration change across the cell diameter which is equal to the minimum required for gradient detection [40]. Recalling (10), we thus obtain

$$\frac{R}{L} \frac{dL}{dx} = (N_A L V)^{-1/2}.$$

In fact, arguing as in [11], the right-hand side in the equation above can be replaced by a more precise estimate, namely,

$$\frac{R}{L} \frac{dL}{dx} = \left(2\pi T D R \left(\frac{N_s}{N_s + \pi a} \right) \left(\frac{K_D L}{K_D + L} \right) \right)^{-1/2}, \quad (59)$$

where D is the ligand diffusion coefficient and T is the total averaging time (cf., for instance, (46)). Solving for dx gives

$$dx = R \left(2\pi T D R \left(\frac{N_s}{N_s + \pi a} \right) \left(\frac{K_D L}{K_D + L} \right) \right)^{1/2} \frac{dL}{L}. \quad (60)$$

Equation (60) has to be supplemented with suitable initial values. A proposal made in [40] is that

$$L = L_{\max} \quad \text{at } x = 0. \quad (61)$$

The maximum guidance distance $x = x_M$ is then defined by

$$L = 0 \quad \text{at } x = x_M. \quad (62)$$

Assuming $L_{\max} \sim 100K_D \gg K_D$, an analysis in [40] yields

$$x_M \sim 1 \text{ cm}, \quad (63)$$

which seems to be a good estimate for the case of neural navigation. Notice that, when rescaled in an appropriate way, (63) corresponds to a distance of ~ 1 km for an organism of the size of a human being.

2.5. Receptor clustering

While the performance predicted by the models described before is fairly good, it has become apparent that cells can do even better. For that reason, the assumptions initially made in [11] were thoroughly revised later, in an attempt to match the experimental facts observed. This led to extensive work on two issues to be considered in our forthcoming sections: the control properties of the intracellular signaling cascade, and the cooperative effects derived from receptor clustering. We shall leave the first from these for the following section, and will concentrate in the second one herein.

It was long since noticed that the nature of the cell membrane allows for lateral mobility of receptors; actually, an estimate on receptors diffusivity was derived as early as in 1975, see [100]. On the other hand, the relevance of multiple attachment to multifunctional ligands (only the monovalent case was considered in [11]) was soon recognized. As it is often the case, theoretical analysis came first (cf. [89–91]) and structural information on the nature of the process was available later (see, for instance, [58,71]).

We shall next describe the early model for receptor clustering proposed by Perelson and De Lisi on [91]. These authors considered the case of reversible binding of bivalent ligands under the assumption (subsequently weakened) that ligands are endowed with two functional units, that may bind to different receptors.

Following [91], let us denote by $L(t)$ the concentration of free ligand in the medium at time t . Suppose that at $t = 0$ all ligand is unbound, and write $L(0) = L_0$, but for later times it can reversibly bind to a receptor with forward (respectively reverse) kinetic constant k_1 (respectively k_{-1}). Let S_0 be the total concentration of receptor sites, present in number n , so that $S_0 = n\bar{S}_0$ where $\bar{S}_0$ is the receptor concentration. Write also $S(t)$ to represent the concentration of free receptor sites at time t . Finally, let $m(t)$ and $M(t)$ respectively denote the concentrations of singly and doubly bound ligands. In order to cross-link two receptors, a free functional group can bind a receptor site located nearby with a rate constant k_2 . If we denote by k_{-2} the kinetic constant for dissociation of a functional group in a doubly bound ligand, one readily arrives at the following system:

$$\frac{dm}{dt} = k_1 L S - k_{-1} m - k_2 m S + 2k_{-2} M, \quad (64)$$

$$\frac{dM}{dt} = k_2 m S - 2k_{-2} M, \quad (65)$$

with initial conditions $m(0) = M(0) = 0$. On the other hand, conservation of ligand and conservation of receptor sites yield the relations

$$L_0 = L(t) + m(t) + M(t), \quad S_0 = S(t) + m(t) + 2M(t). \quad (66)$$

Notice that the model describes the state of the system consisting of cellular receptors and ligand molecules by means of the state of the ligand (free, singly or doubly bound) only. Moreover, knowing that a ligand is singly or doubly bound, does not permit to derive information about the aggregate it is attached to. Furthermore, an equivalent-site hypothesis is also done: no distinction is made between free receptor sites on aggregates of different length, nor between free ligand sites. Finally, intramolecular rearrangements leading to rings of n crossed-linked receptors are also discarded at this stage (although a suitable modification of the model can accommodate such assumption; cf. Section IV in [91]; see also [93]). After discussing the nature of equilibrium solutions to (63)–(65), the question of the distribution of ligand–receptor aggregates in the cell surface is also addressed in [91]. Consider for instance the case of linear chains formed by the interaction of bivalent ligands and bivalent receptors, so that singly bound ligands can only occur at the ends of a chain. Let $c_j(n, t)$ be the concentration of aggregates containing j , $j = 0, 1, 2$, singly bound ligands and n receptors. Then, for $n = 1, 2, 3, \dots$,

$$\begin{aligned} \frac{dc_0(n)}{dt} = & -2k_1 Lc_0(n) + k_{-1}c_1(n) \\ & - 2k_2c_0(n) \left(\sum_{i=1}^{\infty} c_i(i) + 2 \sum_{i=1}^{\infty} c_2(i) \right) \\ & + 2k_2 \sum_{i=1}^{n-1} c_0(i)c_i(n-i) - 2(n-1)k_{-2}c_0(n) \\ & + k_{-2} \left(2 \sum_{i=n+1}^{\infty} c_0(i) + \sum_{i=n+1}^{\infty} c_1(i) \right), \end{aligned} \quad (67)$$

$$\begin{aligned} \frac{dc_1(n)}{dt} = & 2k_1 Lc_0(n) - k_1 Lc_1(n) - k_1c_1(n) + 2k_{-1}c_2(n) \\ & - k_2c_1(n) \left(2 \sum_{i=1}^{\infty} c_0(i) + 2 \sum_{i=1}^{\infty} c_1(i) + 2 \sum_{i=1}^{\infty} c_2(i) \right) \\ & + 4k_2 \sum_{i=1}^{n-1} c_0(i)c_2(n-i) + k_2 \sum_{i=1}^{n-1} c_1(i)c_1(n-i) \\ & - 2k_{-2}(n-1)c_1(n) + 2k_{-2} \sum_{i=n+1}^{\infty} (c_0(i) + c_1(i) + c_2(i)) \end{aligned} \quad (68)$$

and

$$\begin{aligned}
 \frac{dc_2(n)}{dt} = & k_1 Lc_1(n) - 2k_1 c_2(n) - 2k_2 c_2(n) + \left(\sum_{i=1}^{\infty} c_1(i) + 2 \sum_{i=1}^{\infty} c_0(i) \right) \\
 & + 2k_2 \sum_{i=1}^{n-1} c_1(i) c_2(n-i) - 2(n-1)k_{-2} c_2(n) \\
 & + k_{-2} \sum_{i=n+1}^{\infty} (c_1(i) + 2c_2(i)).
 \end{aligned} \tag{69}$$

Equations (66)–(68) make up an infinite system of coupled nonlinear differential equations which is reminiscent of Smoluchowski's system for the coagulation of colloids (cf., for instance, [19]). This last has been widely used to model polymerization and aerosol dynamics (see [36]), and in a simple setting can be formulated as follows. Let $c_n(t)$ denote the concentration at time t of chains consisting of n ($n \geq 1$) identically functional monomers. Assuming monomer aggregation to be irreversible, the c_n 's satisfy

$$\frac{dc_n}{dt} = \frac{1}{2} \sum_{i+j=n} a_{ij} c_i c_j - c_n \sum_{j=1}^{\infty} a_{nj} c_j, \tag{70}$$

where $\{a_{nj}\}$ represent the coagulation coefficients of the process under consideration. If we take all a_{nj} to be equal to a positive constant (Smoluchowski's original assumption), the close relation between (70) on one hand, and (67)–(69) becomes apparent. Actually, for monodisperse initial values (that is, for $c_1(0) = c_0 > 0$ and $c_n(0) = 0$ for $n \geq 2$), system (70) with constant coefficients can be explicitly solved ([19,36]), and the same happens for (67)–(69). More precisely, if we take

$$\begin{aligned}
 c_0(1, 0) &= \frac{S_0}{2}, \\
 c_0(n, 0) &= 0 \quad \text{for } n > 1, \\
 c_1(n, 0) = c_2(n, 0) &= 0 \quad \text{for } n \geq 1,
 \end{aligned}$$

then a combinatorial argument described in [91] reduces the solution of (67)–(69) to that of (64)–(66). More precisely, there holds

$$c_j(n, t) = \frac{S_0}{2} \binom{2}{j} \left(\frac{m}{S_0} \right)^j \left(\frac{2M}{S_0} \right)^{n-1} \left(\frac{S}{S_0} \right)^{2-j}, \quad j = 0, 1, 2. \tag{71}$$

To keep this chapter within reasonable bounds, we shall refrain from discussing the cases of aggregates with rings (for which we refer to Section IV in [91]), ligands with chemically distinct functional groups (considered in [89]), or multivalent ligands, which is analyzed in [90].

At this juncture, the question arises of ascertaining the comparative performances of receptor clustering vs systems of scattered receptors. Actually, we have already noticed that, under the assumptions on receptor nature made in [11], a cluster of receptors would result in a lower efficiency with respect to that of a similar number of well-separated receptor units. The mechanisms by which clustering provides an evolutionary advantage remain largely to be elucidated. Recent research points to amplification in the downstream signaling cascade as (at least part of) an explanation. We shall return to this point in our next section, where properties of intracellular chemical pathways will be considered. We just quote on pass that it has been recently suggested that oligomer formation could actually buffer intracellular signaling against stochastic fluctuations (cf. [1]). In the same work it is also proposed that long linear oligomers increase the range of ligand concentration to which the cell may respond, whereas long closed oligomers seem to favor ligand specificity; see also [2] for related material.

2.6. Signaling pathways and their performance properties

In previous subsection, we have been concerned with the physical mechanisms by which a cell derives information from the concentration of ligands near its surface. We now discuss the properties of signaling pathways. In particular, their ability to amplify signals received as well as to adapt to homogeneous (but largely fluctuating) external ligand concentration will be examined.

It has been already mentioned that chemotactic cells possess a fine sensitivity, that allows them to detect ligand gradients of 1–2% across their surface [5,121]. It is known that these minute differences are internally amplified (even by a factor ~ 55 , cf. [103]). The question naturally arises of understanding the structure of the chemical circuits involved and the output they can provide. To address this issue, we shall begin by following Heinrich, Neel and Rapoport [45] to examine the properties of some simple, although relevant, types of signaling pathways. To that end, let us consider a linear signaling cascade in which stimulation of a receptor leads to consecutive activation of several protein kinases. The eventual output is the phosphorylation of the last kinase, which usually triggers a cellular response (as for instance, activation of a transcription factor). Signaling is inhibited by phosphatases (which dephosphorylate the kinases), and by inactivation of the receptor. To proceed, suppose that each phosphorylation step is described as a reaction between the phosphorylated form X_{i-1} and the nonphosphorylated form $\tilde{X}_i$ of the i th kinase. Assume that the phosphorylation rate is given by $v_{p,i} = \tilde{\alpha}_i X_{i-1} \tilde{X}_i$, and the dephosphorylation rate by $v_{d,i} = \beta_i X_i$, for some kinetic parameters $\tilde{\alpha}_i, \beta_i$. Then the overall process can be represented as follows:

$$\frac{dX_i}{dt} = v_{p,i} - v_{d,i} = \tilde{\alpha}_i X_{i-1} \tilde{X}_i - \beta_i X_i \quad \text{for } 2 \leq i \leq n,$$

where n denotes the total number of subsequently activated kinases, and

$$\frac{dX_1}{dt} = \tilde{\alpha}_1 R(t) \tilde{X}_1 - \beta_1 X_1.$$

Here $R(t)$ is the concentration of activated receptors as a function of time. If we now denote by $C_i = X_i + \tilde{X}_i$, $i \geq 1$, the total amount of kinase i , and set $\alpha_i = \tilde{\alpha}_i C_i$, the equations above read

$$\frac{dX_1}{dt} = \alpha_1 R(t) \left(1 - \frac{X_1}{C_1}\right) - \beta_1 X_1, \quad (72)$$

$$\frac{dX_i}{dt} = \alpha_i X_{i-1} \left(1 - \frac{X_i}{C_i}\right) - \beta_i X_i, \quad 2 \leq i \leq n. \quad (73)$$

To (72), (73) we now add initial values given by

$$X_i(0) = 0, \quad 1 \leq i \leq n, \quad R(0) = R > 0. \quad (74)$$

Moreover, we assume for simplicity that

$$R(t) = R e^{-\lambda t} \quad \text{for some } \lambda > 0. \quad (75)$$

Consideration of any signaling system (and, in particular, the previous one) leads to a number of natural questions. For instance, (i) How fast does the signal reaches its destination? (ii) How long does the signal lasts? and (iii) How can one measure the signal strength? To answer them, the authors of [45] introduce three parameters: the signaling time of the i th kinase, τ_i , given by

$$\tau_i = \frac{T_i}{I_i}, \quad 1 \leq i \leq n, \quad \text{with } I_i = \int_0^\infty X_i(t) dt, \quad T_i = \int_0^\infty t X_i(t) dt, \quad (76)$$

provided that these integrals converge. Notice that τ_i is analogous to the mean value of a statistical distribution. For $1 \leq i \leq n$, the signal duration θ_i is defined as follows:

$$\theta_i^2 = \frac{Q_i}{I_i} - \tau_i^2, \quad \text{where } Q_i = \int_0^\infty t^2 X_i(t) dt, \quad (77)$$

once again, the integral above is assumed to converge, in which case θ_i is similar to the standard deviation of a statistical distribution. Finally, the signal amplitude S_i is defined through the relation

$$S_i = \frac{I_i}{2\theta_i}, \quad (78)$$

so that S_i is the height of a rectangle whose length is $2\theta_i$, and whose area is the same as that enclosed under the curve $X_i(t)$.

Let us briefly recall some of the consequences of the analysis performed in [45]. Consider first the case of weakly activated pathways, for which $X_i \ll C_i$ for any i . Then (73) reduces to

$$\frac{dX_i}{dt} = \alpha_i X_{i-1} - \beta_i X_i, \quad i \geq 2,$$

and the previous parameters can be explicitly computed. More precisely, let τ , θ , S be defined by

$$\tau = \sum_{i=1}^n \tau_i, \quad \theta = \sum_{i=1}^n \theta_i, \quad S = \sum_{i=1}^n S_i.$$

Then there holds

$$\begin{aligned} \tau &= \frac{1}{\lambda} + \sum_{i=1}^n \frac{1}{\beta_i}, \\ \theta^2 &= \frac{1}{\lambda^2} + \sum_{i=1}^n \frac{1}{\beta_i^2}, \\ S &= \frac{R}{2} \prod_{i=1}^n \frac{v_i}{\beta_i} \left(1 + \lambda^2 \sum_{i=1}^n \frac{1}{\beta_i^2} \right)^{-1/2}. \end{aligned} \tag{79}$$

Note that τ , the signaling time through the whole pathway, and θ , the total signal duration, are independent of the kinase rate constants (in other words, they do not depend on the α_i 's). However, the total signal amplitude does depend on all parameters involved in the system.

We say that amplification occurs at the i th step in the cascade if

$$\sigma_i = \frac{S_i}{S_{i-1}} > 1. \tag{80}$$

In the case under consideration, it follows from (79) that (80) is satisfied if

$$\beta_i < \alpha_i \left(1 - \frac{1}{\alpha_i^2 \theta_{i-1}^2} \right)^{1/2}, \tag{81}$$

provided that the quantity within braces is positive. As a matter of fact, it follows from (79) and (81) that longer pathways favor an increase in signaling time, signal duration and amplification in later stages of the signaling cascade.

When weak activation is no longer assumed, one has to deal with the whole system (72), (73). Assuming rapid equilibration; that is, setting $dX_i/dt = 0$ in (73), one obtains

$$X_i = C_i X_{i-1} \left(\frac{\beta_i}{v_i} C_i + X_{i-1} \right)^{-1}. \tag{82}$$

From (80) and (82), one readily sees that amplification occurs if

$$X_{i-1} < C_i \left(1 - \frac{\beta_i}{v_i} \right). \tag{83}$$

A comparison of (81) and (83) shows that signal amplification is less pronounced in this situation. Also an analysis of the case of a permanently activated pathway ($\lambda = 0$ in (75)) for a particular choice of C_i , α_i , β_i shows that amplification only occurs in this example if the activated receptor does not go beyond a threshold value, so that one needs $R < R_*$ for some $R_* > 0$; see Figure 5 in [45] for further details.

A situation also considered in [45] is crosstalk between signaling networks that are simultaneously operating. As an example, consider the case in which a component Y of a second pathway inhibits phosphatase i in the first scheme by changing its kinetic rate. This may be achieved, for instance, by replacing β_i in (73) by

$$\beta_i = \beta_i^0 \left(1 + \frac{Y}{K_i} \right)^{-1},$$

for some positive constants β_i^0 , K_i . It is shown in [45] that crosstalk may have a considerable influence in the case of strong activation, and that it may provide amplification combined with fast and transient signaling.

A final point to be mentioned in this context is that of the stability of signaling networks. In many situations, it is required that random kinase fluctuations should be damped out. However, positive feedback loops are a possible source of instability. Indeed, if we replace (72) by

$$\frac{dX_1}{dt} = (\alpha_1 R + \varepsilon X_n) \left(1 - \frac{X_1}{C_1} \right) - \beta_1 X_1, \quad (84)$$

then instability of the ground state $R = X_i = 0$, $1 \leq i \leq n$, is obtained provided that

$$\beta_1 \beta_2 \cdots \beta_n < \varepsilon \alpha_2 \cdots \alpha_n. \quad (85)$$

On the other hand, from the last formula in (79) it follows that when $\lambda = 0$, the amplification condition at any step in the corresponding circuit reads

$$\beta_1 \beta_2 \cdots \beta_n < \alpha_1 \alpha_2 \cdots \alpha_n. \quad (86)$$

It turns out that sustained amplification can have a destabilizing effect in the presence of feedback loops. We refer to the reader to [45] for further discussion on this and other related topics.

The previous remarks were of a general nature. A different (but complementary) approach consists in analyzing particular situations where amplification is known to occur, in order to unravel the mechanisms that yield such result. A case which has deserved considerable attention is that of the phosphorelay sequence triggered in *E-coli* by the aspartate receptor *Tar*, which eventually connects with the flagellar motors through a pathway involving the *CheA*, *CheY* and *CheZ* proteins [3,58,106]. In particular, in [3] the authors addressed the issue of understanding the reasons for the high gain in the system. This was defined as the change of rotational bias divided by the change in receptor occupancy, and is estimated to be ~ 55 [103].

Out of the several reasons suggested to account for this outstanding performance, the authors of [3] hinted at the indirect activation of many receptors by a single ligand. Receptor clustering, which has been shortly discussed in our previous subsection, has been proposed as a model to enhance sensitivity. However, it has been observed that such mechanism, while seemingly improving sensitivity at low-concentration signals, presents considerable difficulties to simultaneously provide high gain and a wide dynamic range (see [17]).

The model discussed in [3] considers teams of *Tar* receptors (which are of a dimeric nature, cf. [58]) that assemble and disassemble to form teams of one, two or three units. It is assumed that ligand binding destabilizes receptor teams, which subsequently break into smaller units (ligands are not released in that process). A key hypothesis is that only ligand-free threefolds determine kinase activity. The authors remark that their model is able to explain the observed behavior of the kinase activity for a pure receptor under a number of assumptions that, in some cases, could be experimentally tested.

We next turn our attention to adaptation. In engineering terms, any given circuit (possibly representing a chemotactic cell), that can be characterized as producing an output signal ϕ in response to an input signal S is said to possess this property if the output ϕ remains constant when isotropic stimulation S is increased over several orders of magnitude. Bearing chemotaxis always in mind, a question that naturally arises is that of determining (relatively) simple circuits (modules), described by systems of chemical reactions, that display adaptation. Such modules are also expected to be robust. This means that their performance is not significantly altered when parameters in the model undergo large variations (up to some orders of magnitude).

A common feature in many of the models so far derived to account for adaptation is that they make use of activator–inhibitor systems. These last reflect the fact that many examples of biological pattern formation show the interplay of a local, self-enhancing reaction coupled to a long-range antagonistic reaction [39,76]. A typical example is provided by

$$\frac{\partial a}{\partial t} = \frac{\alpha a^2}{h} - \mu a + D_a \frac{\partial^2 a}{\partial x^2}, \quad (87)$$

$$\frac{\partial h}{\partial t} = \delta a^2 - \nu h + D_h \frac{\partial^2 h}{\partial x^2}, \quad (88)$$

where α , μ , δ and ν are kinetic constants, and D_a , D_h denote the respective diffusivities of substances a (activator) and h (inhibitor). Suppose for simplicity that constants α , μ , δ and ν are set equal to one, and assume that activator and inhibitor concentrations are constant in space, so that diffusion effects can be discarded. Then $a = h = 1$ is a solution of the associated kinetic system

$$\begin{aligned} \dot{a} &= \frac{a^2}{h} - a, \\ \dot{h} &= a^2 - h. \end{aligned} \quad (89)$$

According to these equations, if h remains constant and equal to one, and a becomes slightly larger than one, $\dot{a} > 0$ and a will increase further. Actually, should $h = 1$ continue

to hold, the first equation above would then yield finite-time blow-up for $a(t)$. Assume however that there is very rapid equilibration of the inhibitor to a given activator concentration. At the steady state ($\dot{h} = 0$) this gives $h = a^2$, which upon substitution in (89) yields in turn

$$\dot{a} = 1 - a,$$

whence $\dot{a} < 0$ for $a > 1$. In this way, stability of the equilibrium $a = 1$ is achieved. When space inhomogeneities are taken into account, so that a diffusion mass transport sets in, the rapid equilibration of h with respect to a can be achieved by taking $D_h \gg D_a$ in (87), (88), a condition whose relevance in pattern formation was already noticed in Turing's seminal work [114]. In this manner, nonlinear patterns emerge and become stable in (87), (88); see for instance [76] for a detailed discussion on this issue.

As noticed by Meinhardt in [77], this type of model is very convenient to detect minor external concentration differences and convert them into a pronounced intracellular pattern, even when the external signal is subject to random fluctuations. Amplification is thus obtained, and also adaptation, since this effect is rather independent of the value of an external, isotropic stimulus; see Figure 2.

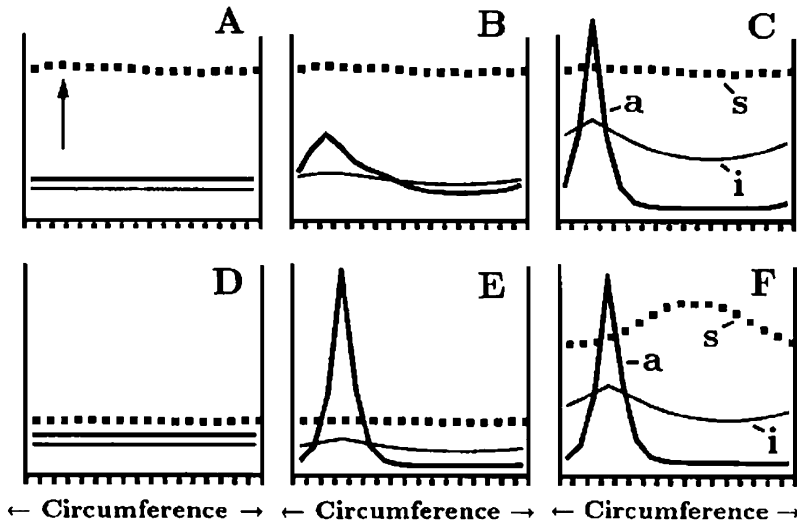


Fig. 2. Detection of minute asymmetries and the problem of reorientation. Assumed is a cell-internal pattern forming reaction consisting of a self-enhancing activator (a) and a long-ranging inhibitor (i). Sufficient for the localization of a strong internal signal is a noisy, slightly asymmetric external signal (s , black squares) that has a stimulating influence on the activator production. Simulations made on the circumference of a circle: the left and right elements are neighbors in reality. Shown is the initial (A), an intermediate (B) and the final stable distribution (C). A strong internal activator maximum appears at the position where the external signal is slightly above average (arrow in A). D, E – As required for path-finding in a graded environment, orientation works also at a much lower absolute level of the external signal. The lower signal concentration is compensated by a lower inhibitor concentration. F – The problem: after an incipient pattern has been formed, even a strong external asymmetry is unable to reorient the pattern. (Reprinted from [77], with permission granted by the Company of Biologists Ltd.)

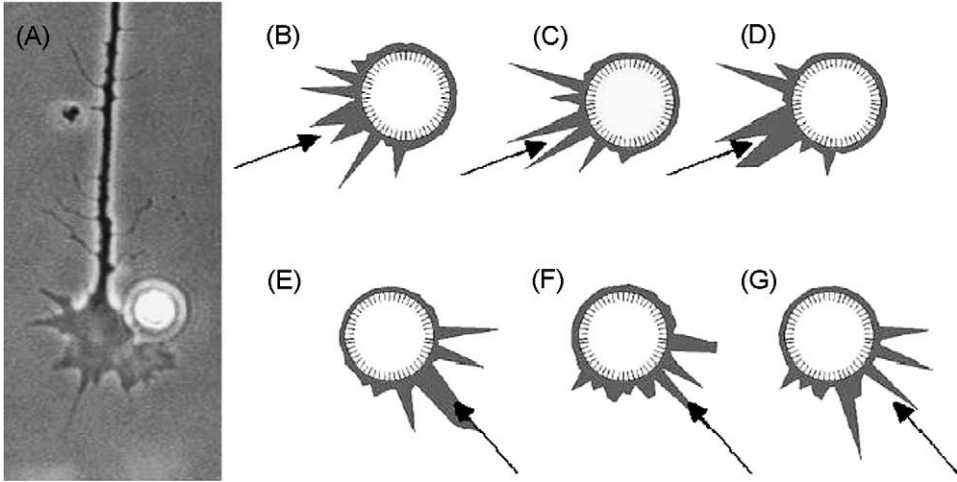


Fig. 3. Orientation of growth cones and chemotactic cells. A – A growth cone of a nerve growing in vitro. B, G – Model. Assumed is an internal pattern-forming system in which the self-enhancing process saturates and in which the activator does not diffuse; shown is only the activation. The distance from the inner circle is a measure for the local activation. The external orienting signal has a positive influence on the internal patterning system of the cell. The concentration difference across the cell is 2%; its orientation is indicated by the arrow. Assumed are max. 1% statistical variations in the cell cortex in the ability to perform the self-activation. B, D – Simulation: somewhat irregular active spots emerge that act as signals to stretch out cell extensions toward the signaling source. Due to their limited half-life caused by a local antagonistic process, they disappear subsequently and new ones emerge instead. E, G – After a change in the orientation of the external signal (arrow), the locations of the temporary signals adapt rapidly to the new direction. Thus, the system is able to detect permanently minute concentration differences (photograph kindly supplied by J. Loschinger). (Reprinted from [78], with permission from Elsevier.)

However, once the intracellular signal (the pattern) is formed, self-stabilization of that pattern is so strong that small external cues (as those actually at work in chemotactic processes) are unlikely to result in a reorientation. To deal with this difficulty, a mechanism was proposed in [77] that consists in including a second antagonistic reaction, to obtain an oscillating activator–inhibitor system. In this way, the cell (nerve growth cones are the example being considered in [77]) proceeds in a cyclic manner from phases where it is highly sensitive to external signals, to periods where weak external inputs are converted into strong internal patterns, that can be used to reorient the cell toward its target. This model is sketched in Figure 3, which is taken from [78].

A particular model where this behavior can be observed is

$$\begin{aligned} \frac{da}{dt} &= \frac{\theta(a^2b^{-1} + \alpha)}{(\beta + c)(1 + \delta a^2)} - \mu a, \\ \frac{db}{dt} &= \nu(a - b), \\ \frac{dc}{dt} &= \omega a - \tau c, \end{aligned} \tag{90}$$

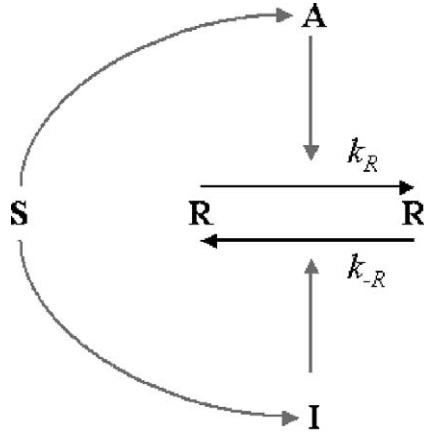


Fig. 4. A schematic representation of system (91)–(93) (adapted from [70]).

for suitable choices of the parameters $\delta, \alpha, \beta, \delta, \mu, \nu, \omega$ and τ therein (see the Appendix in [77]). Actually, in (90) the external signal (and its changing orientation) is lumped in parameter θ , where random fluctuations are also incorporated.

We next discuss a conceptually simple scheme for perfect adaptation to spatially uniform changes in ligand concentration that has been proposed in [70] (see also [62,53]). Its ingredients are as follows. For any given species Z , let us denote by Z_T its total concentration, and by Z^* that of its active form. Suppose that a response element R may go from an active to an inactive state, and that passage from R to R^* (respectively, from R^* to R) is mediated by an activator A (respectively, by an inhibitor I), both of which are produced from an external signal S . We may graphically represent this circuit in Figure 4.

In mathematical terms, the previous scheme can be written in the form

$$\frac{dR^*}{dt} = -k_{-R}IR^* + k_RA(R_T - R^*), \quad (91)$$

$$\frac{dA}{dt} = -k_{-A}A + k'_AS(A_T - A), \quad (92)$$

$$\frac{dI}{dt} = -k_{-I}I + k'_IS(I_T - I). \quad (93)$$

To make the analysis simpler, let us assume that $A_T \gg A$, $I_T \gg I$. Then (91)–(93) reduces to

$$\frac{dR^*}{dt} = -k_{-R}IR^* + k_RAR, \quad (94)$$

$$\frac{dA}{dt} = -k_{-A}A + k_AS, \quad (95)$$

$$\frac{dI}{dt} = -k_{-I}I + k_IS, \quad (96)$$

where $k_a = k'_a A_T$, $k_I = k'_I I_T$, and the various remaining constants in (91)–(93) are kinetic parameters of the process under consideration. At the steady state, we should have

$$\bar{R}^* = \left(\frac{\bar{A}/\bar{I}}{k_D + \bar{A}/\bar{I}} \right) R_T, \quad (97)$$

where $\bar{Z} = \lim_{t \rightarrow \infty} Z(t)$ for any given variable Z , and $k_D = k_{-R}/k_R$ as in (7). Note that (95)–(96) yield, at the steady state

$$\bar{I} = k_1 \bar{S}, \quad k_1 = \frac{k_I}{k_{-I}}, \quad \bar{A} = k_2 \bar{S}, \quad k_2 = \frac{k_A}{k_{-A}},$$

and therefore (97) corresponds to perfect adaptation: the value of $\bar{R}^*$ provided in that formula remains unchanged when $S(t)$ is replaced by $\alpha S(t)$ for any $\alpha > 0$.

The kinetic system (94)–(96) has some drawbacks when considered as a building block for a model that should provide spatial sensing with high gain. For instance, suppose that an external source varies linearly along the length of the cell, so that $S(x) = c_0 + c_1 x$ for some constants c_0, c_1 , where x is a normalized distance measured, say, along a cell diameter. Assume for simplicity that the activator A does not diffuse but the inhibitor I does so according to the equation

$$\frac{\partial I}{\partial t} = -k_{-I} I + k_I S + D \Delta I,$$

which should replace (96) in system (94)–(96). An analysis as that described in [62] shows that

$$\bar{R}^*(x) \sim \frac{1}{1 + f(x)} \quad \text{for some } f(x) \geq 0,$$

so that this model does not provide gain in the difference of activity between the front and the rear of the cell with respect of that of the external signal. A way to remedy this situation consists in increasing the complexity of the kinetic scheme under consideration. This may be achieved, for instance, by replacing the process depicted in Figure 4 by that in Figure 5.

The onset of asymmetry in a chemotactic cell after a rise in external signal is often the first noticeable step in the subsequent directional sensing process. This issue has been addressed in [98] in the context of studies conducted on the slime mold *Dictyostelium discoideum* (*Dd*). The situation succinctly described in [98] is as follows. An applied cAMP signal, way larger than that required to trigger a response, is applied to a medium with *Dd* cells. This signal will quickly diffuse around any individual cell, from front to back. Since cAMP receptors are thought to be uniformly distributed over the cell's membrane, it is natural to assume that an inhibitory mechanism should appear that suppresses responses at the cell's back (measured with respect to the cAMP source introduced). Once an initial asymmetry is established, it should be amplified (and stabilized) by means of appropriate mechanisms.

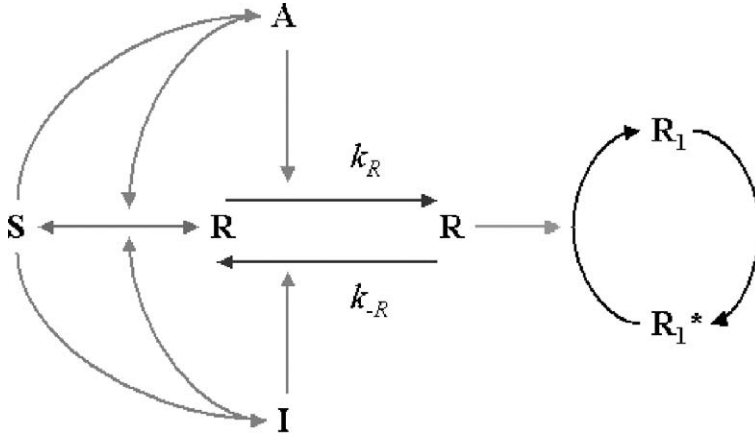


Fig. 5. A kinetic system possessing perfect adaptation that may also provide high gain in the second response element R_1 for a suitable choice of the parameters describing kinetic rates. See the Appendix in [70] for details. (Adapted from [70].)

In [98] a dynamical model is introduced to describe the setting of a rapid initial asymmetry as that described above. This is done as follows. First, a three-state characterization of the membrane is introduced: quiescent (with density ρ_q), activated (density ρ_a) and inhibited (density ρ_i). Since the total density has to be conserved, $\rho_q + \rho_a + \rho_i = 1$. The equations proposed to describe the membrane state are

$$\begin{aligned}\frac{\partial \rho_q}{\partial t} &= -\alpha c \rho_q + \beta_f \rho_i - \beta_i g \rho_q, \\ \frac{\partial \rho_a}{\partial t} &= \alpha c \rho_q - \delta \rho_a, \\ \frac{\partial \rho_i}{\partial t} &= -\beta_f \rho_i + \beta_r g \rho_q + \delta \rho_a,\end{aligned}$$

which satisfy the previous conservation requirement. Here, α , β_f , β_i , δ and β_r are kinetic constants and c and g respectively denote the concentrations at the membrane of extracellular and intracellular cAMP. These concentrations in turn satisfy

$$\frac{\partial c}{\partial t} = D_c \Delta c - v_c c,$$

at the extracellular space, and

$$\frac{\partial g}{\partial t} = D_g \Delta g - v_g g,$$

at the intracellular one; v_c and g_c represent decay terms. It is also assumed that a no-flux condition is satisfied at the boundary

$$D_g \nabla g \cdot \mathbf{n} = D_c \nabla c \cdot \mathbf{n} = 0,$$

where $\mathbf{n}$ denotes the normal at the membrane. The simulations performed in [98] for a suitable choice of parameters (and for equal diffusivities: $D_c = D_g = 2.5 \times 10^{-6} \text{ cm}^2/\text{s}$) show that asymmetry, characterized in terms of the values of ρ_a , ρ_i is rapidly (less than 1 second) established in the cell (see Figures 2 and 3 in [98]).

We have already remarked on the meaning of adaptation from an engineering point of view. Further discussion on this subject is contained in [120], where the authors recall that there are two ways of constructing systems exhibiting perfect adaptation. The first possibility requires, as observed in [120], fine-tuning the parameters in the corresponding model, an approach consistently followed in the works described so far. A second alternative consists in designing specific structures that create such property inherently. In this vein, adaptation may be viewed as a solution for a common problem in engineering, namely that of designing systems that (quickly) converge toward a specific steady-state output.

A standard solution to this problem is integral feedback control. In its simplest setting, this process can be described by the equations

$$\dot{x} = y, \tag{98}$$

$$y = y_1 - y_0 = k(u - x) - y_0, \quad k > 0. \tag{99}$$

In (98), (99) a process is represented that takes u as an input and produces the output y_1 . This process is characterized by constant k in (99). We denote by y the difference between y_1 and the steady-state output y_0 . This represents the system error, whose time integral x is fed back into the system, with the aim of obtaining the desired result

$$y(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty. \tag{100}$$

As recalled in [120], for linear systems a necessary and sufficient condition for robust asymptotic tracking is that the system had integral feedback, as that described by (98).

As observed in [120], an interesting example where these ideas can be applied is the case considered in [6]. These authors derived a two-state (active or inactive) model of a receptor complex, constituted by a receptor and *CheA* and *CheW* proteins, in bacterial chemotaxis. The system output was the concentration of active receptor complexes. A remarkable fact stressed in [6] is that perfect adaptation is achieved as an intrinsic property of the signaling network considered, independently of the kinetic parameters involved. It was then shown in [120] (supplementary material) that a system of differential equations can be written for the biochemical network described in [6] such that, after suitable manipulation, an equation characteristic of integral control is derived. The activity of the system is then shown to asymptotically converge to a fixed steady-state value (cf. (1) in [120]). As the authors of [120] point out, knowing that integral control underlies the robustness of perfect adaptation in the model designed in [6] has significant implications. In particular, it allows for assessing the relevance of any of the several assumptions made in that model. A discussion

on this subject can be found in [120], and the reader is referred to that work for further details on that issue.

3. Some mathematical problems arising from the study of *Dictyostelium discoideum*

It has been already recalled at the Introduction of this chapter that chemotaxis was proposed as a hypothesis in connection with the process of neural wiring in the field of Neuroscience. This last has proved to be a fertile ground in Biology during the XX century, and has exerted a considerable influence in the evolution of other disciplines as Mathematics and Physics. This situation is likely to continue, even at an increased pace, during the foreseeable future. A discussion on current and future research directions at the interface between Mathematics, Neuroscience and Physics can be found in the monograph [116] and references therein.

At a different level of complexity, a relevant role in the blossom of Developmental Biology started at the beginning of the last century has been played by the study of animal models, particularly (but by no means exclusively) those made on the bacteria *E-coli* and the slime mold *Dictyostelium discoideum* (*Dd*). In both cases chemotaxis is a central topic when describing developmental and social life properties of such microorganisms. The reader is referred to [10] for a recent description of current understanding of *E-coli* from a multidisciplinary point of view.

In this section we shall focus on *Dd*, and will discuss some of the mathematical approaches that have been proposed to deal with quantitative problems motivated by the study of such organism. More precisely, the plan of this section is as follows. In Section 3.1 we shortly review some basic facts concerning the biology of *Dd*. Particular attention will be paid to the role played by a chemical messenger (cAMP) in starvation-induced aggregation into some condensation centers, a remarkable feature of *Dd* colonies which has triggered large attention in the biology community. We then focus in Section 3.2 into a particular set of differential equations (the so-called Keller–Segel system) which was initially proposed as a model to describe early stages of aggregation in *Dd* cell cultures, and has been extensively studied by mathematicians since, due to the nontrivial structure displayed by their solutions. In that paragraph the emphasis is therefore on the mathematics, keeping however an eye on the biological motivations. As everywhere in this work, the style will be descriptive, and reference will be made to appropriate articles for details on the underlying mathematical arguments.

While the Keller–Segel (KS) model described in Section 3.2 is of a macroscopic nature, the problem of relating observable macroscopic behavior in *Dd* cultures to individual cell properties has sparked interest in deriving KS equations from microscopic considerations, a multiscale problem of considerable importance, which will be dealt with in Section 3.3. It should be stressed that the considerations to be recalled there are by no means limited to the case of *Dictyostelium*, but rather point at the deep, general question of relating macroscopic properties of organic ensembles (swarms, tissues, organs, . . .) to the individual signal exchange and transduction pathways of their members.

We then conclude this section with a discussion on some aspects of pattern formation in *Dd* colonies, particularly in the case of monolayer cultures that can be represented by

two-dimensional (in space) domains. Specifically, we concentrate on early stages of aggregation, and in particular in the occurrence of target and spiral waves and the transition from prevalence of one type of pattern to the other. In mathematical terms, this leads to the question of characterizing reaction–diffusion systems that are able to exhibit that type of traveling waves, a subject which is quickly addressed therein. Once again, understanding the observed dynamic transition from a target- to spiral-dominated scenario seems to call for a multiscale approach which is only at its beginning yet.

3.1. *The social life of Dictyostelium discoideum in a nutshell*

There is a number of recent reviews on *Dd* where the biology of this species is described in detail (cf., for instance, [23,34,48,115]). For definiteness, we shall recall below only a few aspects of that topic which are particularly relevant as background for the mathematical models to be mentioned later.

The basic features of individual and collective behavior in *Dd* colonies are best summarized in the words of John Bonner ([16], p. 62) as follows:

... Cellular slime molds are soil amoebae. They feed as separate individuals on bacteria, and after they have finished the food supply, they stream together to central collection points to form a multicellular individual of thousands of cells....

In fact the cells migration toward aggregation centers (the “central collection points” in the quotation above) is mediated by a chemical compound (adenosine 3',5'-cyclic monophosphate, cAMP; cf. [59]) which is produced by cells at the aggregation centers in a pulsatile way upon starvation, and spreads by diffusion [75]. Moreover, *Dd* cells are able to relay the cAMP signal received, thus keeping a cAMP flow in the medium that, when observed through darkfield microscopy, gives raise to fields of circular and spiral waves (for a review, see [48]). This cAMP-mediated chemotactic migration eventually results in the formation of mounds or condensates which, most remarkably, have a rather constant size [99].

Once mounds have been formed, differentiation of mound cells into two cell types (prespore and prestalk) begins. Prestalk cells then migrate to the upper part of the mound to form a tip, whereas prespore cells remain in the lower side of the mound. The object thus formed then elongates to produce a finger-like structure. Subsequent development can unfold in two different ways, as recalled in Figure 6. The first case occurs under favorable environmental conditions, and consists in finger development at the place where it was formed. Prestalk cells, that were situated at the top of the structure, migrate through the prespore cells in the direction of the substratum. Meanwhile, these cells differentiate into stalk cells and die at the end of the process. The growth of the stalk, where prestalk cells are continuously added at its top, goes in parallel with the upward movement of prespore cells, that rise from the substratum to eventually form a ball of cells on top of the stalk (the sorus). Prespore cells in the sorus eventually become spores surrounded by a hard shell, that remain viable for weeks. Germination of spores leads to a new cell cycle for the resulting amoebae.

If conditions are unfavorable, a longer developmental program can be selected. Then the finger structure falls on the substrate and a migratory structure (the slug) is formed,

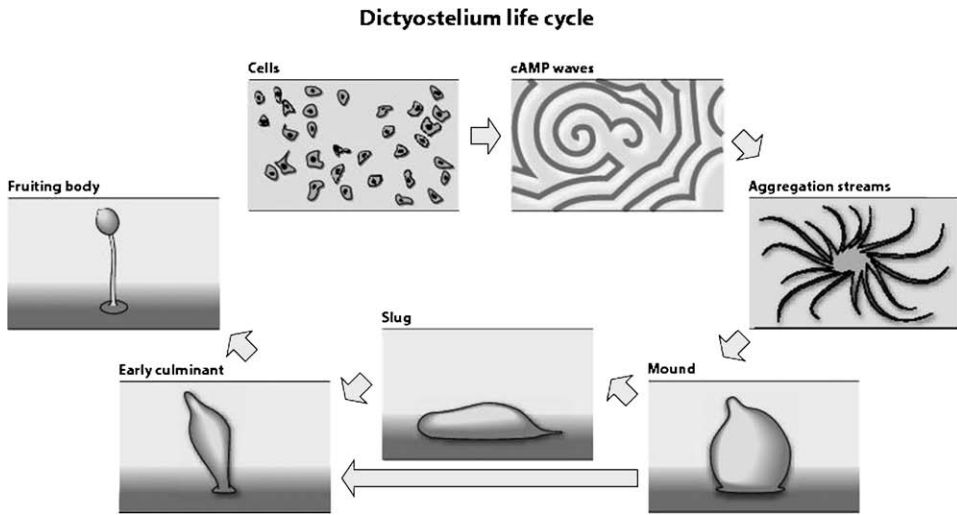


Fig. 6. Evolution of *Dd* cultures after starvation. (Reproduced with permission from [48].)

where prestalk cells are located at the front and prespore cells at the back. Slugs can move toward places more suitable to complete development. When this happens, the slug stops, rounds up and migration of prestalk and prespore cells in opposite direction occurs as in the previous case, after which culmination takes place in the same manner.

3.2. Early aggregation stages: the Keller–Segel model

Out of the many mathematical models that have been proposed to deal with particular aspects of chemotaxis, that proposed by Keller and Segel in 1970 (cf. [57]) has received particular attention. There are a number of features on it that can possibly explain the interest it has raised among the mathematical community (witnessed, for instance, by the thorough survey [51] and the monograph [111]). For instance, the model has a very simple structure, reflecting the fact that the underlying hypotheses are reduced to a bare minimum. Moreover, the mathematical analysis of their solutions is nontrivial, and has led to developments of considerable interest. However, the model has proved to be fairly less popular among biologists interested in chemotaxis. Some thoughts on that matter, due to one of the authors of [57], can be found in [56], Chapter 1. See also [41] for a number of interesting remarks on the role of modeling in Biology.

The systems considered in [57] were intended to describe the early aggregation properties of the slime mold *Dictyostelium discoideum*. In particular, the question of the formation of condensates where *Dd* cells gather upon starvation, was paramount there. The authors' approach was neatly explained by them at the Introduction of their work:

... By analogy with many problems in the physical world, aggregation is viewed as a breakdown of stability caused by intrinsic changes in the basic parameters that characterize the system ... ,

a statement which bears resemblance with that made by Turing in [114]:

... A system of chemical substances... although it may originally be quite homogeneous, may later develop a pattern or structure due to an instability of the homogeneous equilibrium, which is triggered off by random disturbances

Let us briefly recall the analysis done in [57] for the simplest model considered there. Let $u(x, y, t)$, $v(x, y, t)$ respectively denote the concentrations of amoebae and cAMP in a two-dimensional medium, which can be thought of as an approximate setting for monolayer *Dd* culture in Petri dishes. Then u and v are required to satisfy

$$\frac{\partial u}{\partial t} = -\nabla(D_1 \nabla v) + \nabla(D_2 \nabla u), \quad (101)$$

$$\frac{\partial v}{\partial t} = D_v \Delta v - k(v)v + uf(v), \quad (102)$$

where

$$k(v) = \frac{k_1}{1 + kv}, \quad (103)$$

for some constants $k_1 > 0$ and $k > 0$. In this system, the right-hand side in (101) includes the contributions to time-change of u due to convective motion induced by cAMP and diffusion respectively. In particular, $D_1 = D_1(u, v) > 0$ represents a measure of the influence of the cAMP gradient on the flow of amoebae, and $D_2 = D_2(u, v) > 0$ is a diffusion coefficient corresponding to random motion of *Dd* cells. The right-hand side in (102) has a diffusion term for cAMP (with diffusion coefficient $D_v > 0$), a decay term for that chemical (with kinetic parameter $k(v)$ given in (103)) and a source term, which corresponds to the assumption that cAMP is produced by the cells themselves. In this last case, the kinetic parameter is given by $f(v)$ in (102) and is of a general form. To make up for a well-posed mathematical problem, equations (101)–(103) need to be supplemented with initial values and boundary conditions. When the problem is considered in a bounded domain $\Omega \subset \mathbb{R}^2$, these last are usually taken to be of no-flux type, namely

$$\frac{\partial u}{\partial n} = \frac{\partial v}{\partial n} = 0 \quad \text{at } \partial\Omega, \quad (104)$$

where $\partial\Omega$ denotes the boundary of Ω , and n stands for the (outer) normal at $\partial\Omega$.

A particular type of solution to (101)–(103) are equilibria, defined by

$$u = u_0, \quad v = v_0 \quad \text{with } u_0 f(v_0) = k(v_0)u_0, \quad (105)$$

provided that the last equation in (105) admits constant roots. A relevant contribution in [57] consists in obtaining conditions under which the steady state (u_0, v_0) in (105) is unstable under perturbations. This instability is thus seen as the first step toward the formation of a condensate. The previous goal is achieved by means of a classical linear stability argument, that is summarized as follows. Set

$$u = u_0 + \bar{u}(x, y, t), \quad v = v_0 + \bar{v}(x, y, t), \quad |\bar{u}| \ll u_0, |\bar{v}| \ll v_0. \quad (106)$$

Plugging (106) into (101), (102) and neglecting higher-order terms, $\bar{u}$ and $\bar{v}$ are shown to satisfy the linear system

$$\frac{\partial \bar{u}}{\partial t} = -D_1(u_0, v_0)\Delta \bar{v} + D_2(u_0, v_0)\Delta \bar{u}, \quad (107)$$

$$\frac{\partial \bar{v}}{\partial t} = D_v\Delta \bar{v} - \bar{k}\bar{v} + u_0 f'(v_0)\bar{v} + f(v_0)\bar{u}, \quad (108)$$

where $\bar{k} = k(v_0) + v_0 k'(v_0)$. Assuming for simplicity that (107), (108) is satisfied in the whole plane $\mathbb{R}^2$, and trying there

$$\bar{u} = \hat{u} \cos(q_1 x + q_2 y) e^{\sigma t}, \quad \bar{v} = \hat{v} \cos(q_1 x + q_2 y) e^{\sigma t}, \quad (109)$$

for some constants $\hat{u}$, $\hat{v}$, q_1 , q_2 and σ , one readily obtains

$$(F - \sigma)\hat{v} + f(v_0)\hat{u} = 0, \quad D_1 q^2 \hat{v} - (D_2 q^2 + \sigma)\hat{u} = 0,$$

where $q^2 = q_1^2 + q_2^2$ and $F \equiv f'(v_0)u_0 - \bar{k} - q^2 D_2$, and D_1 , D_2 are evaluated at (u_0, v_0) . The previous system has a nontrivial solution provided that

$$\sigma^2 - \sigma(F - q^2 D_2) - (q^2 f(v_0)D_1 + q^2 D_2 F) = 0. \quad (110)$$

Analysis of this quadratic equation shows that the corresponding roots are real, and the condition for obtaining values $\sigma > 0$ is that

$$\frac{D_1 v_0}{D_2 u_0} + \frac{u_0 f'(v_0)}{\bar{k}} > 1. \quad (111)$$

Summing these results up, we have seen that (111) is required for σ to be positive in (109), so that the corresponding oscillatory perturbation of the steady state tends to increase in amplitude for small times. However, this argument does not imply that such a perturbation will increase for ever. Indeed, as $\bar{u}$ and $\bar{v}$ in (109) grow larger, the higher-order terms which were discarded in (107), (108) become relevant, and the previous argument no longer applies. Thus a fully nonlinear analysis for intermediate times is required in order to examine the formation of condensates.

A first step in that direction was later provided by Nanjundiah (cf. [82]). He considered a simplified version of (101), (102), namely

$$\frac{\partial u}{\partial t} = \nabla(D_u \nabla u - \chi u \nabla v), \quad (112)$$

$$\frac{\partial v}{\partial t} = D_v \Delta v + A u - B v \quad (113)$$

(the so-called gradient proportional chemotaxis in [82]), where the diffusion coefficients D_u and D_v , and parameters χ , A and B are positive constants. Incidentally, this system is commonly referred to as the Keller–Segel system in the mathematical literature.

Nanjundiah proposed that condensate formation in *Dd* monolayer cultures would correspond to blow-up in finite time T for (112), (113), in such a manner that a Dirac-delta type singularity will develop as the blow-up time is approached:

$$u(x, t) \rightarrow M\delta(x - x_0) \quad \text{for some } M > 0 \text{ and some } x_0 \text{ as } t \rightarrow T, \quad (114)$$

a property commonly referred to as the onset of a chemotactic collapse. We remark on pass that equations (112), (113) can be made dimensionless upon the change of variables:

$$u \rightarrow u^* = \frac{A\chi}{BD_u}u, \quad v \rightarrow v^* = \frac{\chi}{D_u}v, \quad r \rightarrow r\sqrt{\frac{B}{C_v}}, \quad t \rightarrow Bt, \quad (115)$$

which transforms (112), (113) into

$$\frac{\partial u^*}{\partial t} = D\nabla(\nabla u^* - u^*\nabla v^*), \quad D = \frac{D_u}{D_v}, \quad (116)$$

$$\frac{\partial v^*}{\partial t} = \Delta v^* + u^* - v^*. \quad (117)$$

Notice that in this case the instability condition (111) reads

$$u^*(x, 0) > 1, \quad (118)$$

which suggests that the uniform distribution is unstable above a critical concentration of amoebae. Further analysis was later done in [22], where it was shown that no blow-up occurs in finite time for (112), (113) in space dimension $N = 1$, and in [21] where an asymptotic argument was presented to show that, in space dimension $N = 2$ chemotactic collapse will occur if the cell's density goes above some threshold. More precisely, the statement in [21] goes as follows. Consider radial solutions of (112), (113) in a ball $B_L = \{(x, y): x^2 + y^2 < L^2\}$, and introduce new variables given by

$$\tilde{u} = \frac{A\chi L^2}{D_u D_v}u, \quad \tilde{v} = \frac{\chi}{D_u}v, \quad \rho = \frac{r}{L}, \quad \tau = \frac{D_u t}{L^2}. \quad (119)$$

In these new dimensionless variables, (112), (113) is changed into

$$\begin{aligned} \frac{\partial \tilde{u}}{\partial \tau} &= \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \left(\frac{\partial \tilde{u}}{\partial \rho} - \tilde{u} \frac{\partial \tilde{v}}{\partial \rho} \right) \right), \\ \delta_1 \frac{\partial \tilde{v}}{\partial \tau} &= \tilde{u} - \delta_2 \tilde{v} + \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \tilde{v}}{\partial \rho} \right), \end{aligned}$$

where now $0 < \rho < 1$, and

$$\delta_1 = \frac{D_u}{D_v}, \quad \delta_2 = \frac{BL^2}{D_v}. \quad (120)$$

Then, in the limit where $0 < \delta_1 < 1$ and $0 < \delta_2 < 1$, a formal argument is provided in [21] that shows that for chemotactic collapse to occur, one needs

$$\int_0^1 \rho \tilde{u}(\rho, 0) d\rho > 4.$$

When written in terms of the original variables in (112), (113), this condition reads

$$M_0 \equiv \int_{B_L} u_0(x, y) dx dy > \frac{8\pi D_u D_v}{A\chi}. \quad (121)$$

The previous argument hints at a threshold result for chemotactic collapse when $N = 2$. Namely, when (112), (113) are considered with no-flux conditions (104), so that the cell population is preserved,

$$\int_{B_L} u(x, y, t) dx dy = \int_{B_L} u_0(x, y) dx dy,$$

as long as the solution exists, we expect that the value

$$M^* = \frac{8\pi D_u D_v}{A\chi} \quad (122)$$

will play a critical role, in that solutions should be global in time for $M_0 < M^*$ (M_0 being as in (121)) and condensates of Dirac-delta type should form when $M_0 > M^*$. As it turns out, this happens to be the case when radial solutions are considered. Moreover, whenever blow-up occurs at an interior point for (112), (113), it has to correspond to chemotactic collapse at that point (cf. (114)) with $M = M^*$. We refer to the reader to [51] for a detailed account of the precise statements concerning global existence and blow-up for (112), (113) as well as for exhaustive references (up to 2003) of the works where such results were obtained. Among these last we should merely mention here the articles [12,49,54,80] and [81] as illustrative of the different techniques developed to derive the results just sketched.

At this juncture, it is worth to mention that chemotactic collapse (as described in (114)) is known to be the only possible type of singularity formation in space dimension $N = 2$ for system (112). However, when $N = 3$ a different type of hydrodynamic collapse has been shown to occur in [47] for a simplified version of (112) in the whole space.

We conclude this section with some remarks on recent developments motivated by (or related to) systems akin to (112)–(113). It goes without saying that reference to the works that follow is far from being complete, and the reader will find additional information by consulting the articles mentioned. To begin with, the occurrence of singularities for models similar to (112), (113) has been recently examined in [13,14] in the critical case when $M_0 = M^*$ (cf. (121), (122)). Global existence results for equations of Keller–Segel type have been derived in [24], whereas an outline of PDE models (parabolic and hyperbolic) to describe chemotaxis can be found in [92]. Recently, systems of Keller–Segel type with nonlinear diffusivities have been considered by a number of authors, among which we

should mention [109,110] and [72]. Back to the original KS model, where blow-up is taken as a fingerprint for aggregation in mounds, a natural question consists in ascertaining in which sense (if any) solutions can be continued after condensate formation. This issue has been addressed in [117,118].

3.3. The Keller–Segel model revisited: from micro to macro

The system (101)–(103), as well as the variants thereof considered in [57], were obtained by means of macroscopic (and largely heuristical) considerations. It has been shown, however, that equations of this type can be derived starting from microscopic models, much in the same way as the linear diffusion equation can be arrived at as a macroscopic limit for random walks (cf., for instance, [9]). We next recall various manners in which macroscopic chemotaxis equations can be obtained from microscopic models.

(I) *Chemotaxis and biased random walks.* Next we shall borrow from [86] and [87], and quickly review how to derive a chemotaxis model from a master equation for a continuous in time, discrete in space random walk on a one-dimensional lattice. To this end, let us define $u_i(t)$ as the probability of a walker to be at an (integer) point i at time t , starting from $i = 0$ at $t = 0$. Assume now that the random walk evolves according to the equation

$$\frac{\partial u_i}{\partial t} = T_{i-1}^+ u_{i-1} + T_{i+1}^- u_{i+1} - (T_i^+ + T_i^-) u_i, \quad (123)$$

where $T_i^\pm(\cdot)$ denote the transitional probabilities per unit time of a one-step jump to $i \pm 1$. To account for chemotaxis, a spatial bias is introduced, so that one writes $T_i^\pm = T_i^\pm(v)$, where v denotes the chemical concentration in the lattice. If we assume that cells can detect a local gradient, we may write

$$T_i^\pm = \alpha + \beta(\tau(v_{i\pm 1}) - \tau(v_i)), \quad (124)$$

where α and β are positive parameters, and $\tau(\cdot)$ is a function which depends on the particular mechanism for signal detection being considered. Plugging (124) into (123), one readily obtains

$$\begin{aligned} \frac{\partial u_i}{\partial t} = & \alpha(u_{i+1} - 2u_i + u_{i-1}) \\ & - \beta((u_{i+1} + u_i)(\tau(v_{i+1}) - \tau(v_i)) - (u_i + u_{i-1})(\tau(v_i) - \tau(v_{i-1}))). \end{aligned}$$

We then set $x = ih$, and consider x as a continuous variable. We also postulate that the transitional probabilities change according to the scaling $T_h^\pm = (k/h^2)T^\pm$ for some $k > 0$. On extending the definition of u_i in a corresponding manner, and neglecting terms of order $O(h^2)$, one eventually arrives at

$$\frac{\partial u}{\partial t} = D_u \frac{\partial^2 u}{\partial x^2} - \frac{\partial}{\partial x} \left(u \chi(v) \frac{\partial v}{\partial x} \right), \quad (125)$$

where

$$D_u = k\alpha, \quad \chi(v) = 2k\beta \frac{d\tau(v)}{dv}$$

(compare with (101)). Notice that no mechanism for generation of the chemical species has been proposed as yet. A simple way of addressing this issue consists in coupling (125) with a phenomenological equation

$$\frac{\partial v}{\partial t} = D_v \frac{\partial^2 v}{\partial x^2} + g(u, v) \quad (126)$$

for some $D_v > 0$ and some kinetic function $g(u, v)$ (compare with (102)). We remark on pass that it is easy to incorporate counting mechanisms in the model that limit the size of the aggregates. For instance, arguing as in [87] we may replace (124) by

$$T_i^\pm = q(u_{i\pm 1})(\alpha + \beta(\tau(v_{i\pm 1}) - \tau(v_i))), \quad (127)$$

which amounts to assume that the probability of jumping into a neighboring site depends on the space actually available there. For instance, a possible choice for $q(u)$ is

$$q(u) = 1 - \frac{u}{u_{\max}} \quad \text{for } 0 < u < u_{\max}.$$

When we repeat the previous argument with (124) replaced by (127), we readily see that (125) has to be replaced by

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left(D_u (q(u) - q'(u)u) \frac{\partial u}{\partial x} - q(u)u\chi(v) \frac{\partial v}{\partial x} \right). \quad (128)$$

We point out that the existence of attractors for systems of the type (126), (128) has recently been proved in [119].

In the context of our current discussion on global existence vs blow-up for Keller–Segel type systems, it is worth remarking on work conducted on a related model, which has been used to describe vasculogenesis (cf. [4,37]) and reads as follows:

$$\begin{aligned} \frac{\partial n}{\partial t} + \nabla(n \cdot v) &= 0, \\ \frac{\partial v}{\partial t} + v \nabla v &= \mu \nabla c - \beta v - \nabla g(n), \\ \frac{\partial c}{\partial t} &= D \Delta c + \alpha n - \frac{c}{\tau}, \end{aligned} \quad (129)$$

where n and c denote the concentrations of cells and chemoattractant respectively, and v stands for the cell velocity. The equations above take into account cell migration and

chemotaxis (μ measuring the strength of the cell response), friction with substratum (parameter β) and pressure exerted by cells (represented by $g(n)$). As to D , α and τ they respectively represent the diffusion coefficient of the chemical, its production rate and a characteristic degradation time. Neglecting persistence (that is, the inertial term) in the second equation, one obtains

$$0 = \mu \nabla c - \beta v - \nabla g(n)$$

or

$$v = \chi \nabla c - \nabla h(n),$$

with $\chi = \mu/\beta$, $h(n) = 1/\beta g(n)$. Plugging the equation just derived for v into the mass balance equation above, one eventually obtains

$$\begin{aligned} \frac{\partial n}{\partial t} &= \nabla (n \nabla h(n) - \chi n \nabla c), \\ \frac{\partial c}{\partial t} &= D \Delta c + \alpha n - \frac{c}{\tau}, \end{aligned}$$

which is of the Keller–Segel type previously considered. It has been shown in [60] that for functions $h(n)$ that increase fast enough blow-up cannot occur. Previously, a stability analysis of homogeneous solutions to the full system (129) was done in [61]. In particular it was shown that the effect of pressure controls possible instabilities at low densities, thus preventing initiation of blow-up.

(II) *Chemotaxis and velocity jump processes.* Velocity-jump processes are a particular type of stochastic processes, which roughly speaking can be described as consisting in a sequence of runs, separated by reorientations, where a new velocity is chosen (cf. [83] for a detailed discussion). In mathematical terms, they lead to the consideration of Boltzmann equations, namely

$$\frac{\partial p}{\partial t} + \nabla_x v p = -\lambda p + \lambda \int T(v, v') p(x, t, v') dv'. \quad (130)$$

Here $p(x, v, t)$ denotes the density function in a $2n$ -dimensional space with coordinates $x \in \mathbb{R}^n$ (representing the position of an individual) and $v \in \mathbb{R}^n$ (describing its velocity). In this way, $p(x, v, t) dx dv$ gives the number density of individuals whose position is located between $(x, x + dx)$, and whose velocity lies within $(v, v + dv)$. In (130) it is assumed that the random velocity changes follow a Poisson process of intensity λ , so that λ^{-1} is the mean run length time between random choices of direction. On the other hand, the kernel $T(v, v')$ represents the probability of a velocity change v' to v , and $\int T(v, v') dv = 1$ for any v .

The derivation of systems as (101), (102) from equations of the type (130) has been discussed by several authors (see, for instance, [18,50,84], etc.). Here we shall remark

on a suitable extension of the approach in [50,84] which has been introduced by Erban and Othmer [32,33] and where dependence on internal cell kinetics, as that recalled in Section 2.6, is taken into account; see also [7] and [8] for a detailed discussion on related multiscale topics.

The authors of [32], [33] assume that the density of the biological population being considered $p(x, v, y, t)$ depends also on an internal state $y \in \mathbb{R}^m$, $m \geq 1$, which is supposed to follow the kinetics given by a system of the form

$$\frac{dy}{dt} = f(y, S), \quad (131)$$

where $y = (y_1, \dots, y_m)$ denotes the internal variables or species, $S(x, t) = (S_1, \dots, S_d) \in \mathbb{R}^d$ corresponds to the chemical cues acting in the environment, and f represents the precise internal dynamics of the process (for a particular choice of f , see [107]). In addition, the mean run length is also assumed to depend on y . In this way, (130) is to be replaced by

$$\frac{\partial p}{\partial t} + \nabla_x v p + \nabla_y f p = -\lambda(y) p + \int \lambda(y) T(v, v', y) p(x, t, v', y) dv'. \quad (132)$$

A perturbative analysis performed in [32,33] shows that, to the lowest order, the cell variable

$$n(x, t) = \iint p(x, t, v, z) dv dz$$

satisfies the equation

$$\frac{\partial n}{\partial t} = \nabla(D \nabla n - n \chi(S) \nabla S), \quad (133)$$

provided that a number of assumptions are made. In particular, the internal dynamics is assumed to be described by two variables y_1, y_2 which represent excitatory and inhibitory mechanisms respectively. Furthermore, $S = S(x)$ is taken to be independent of time. In equation (133), D and χ are respectively the diffusion and chemosensitivity tensors, so that an equation of type (101) is recovered in the isotropic case.

(III) *Chemotaxis equations as limits of stochastic many-particle systems.* The next approach to be succinctly reviewed here is due to Stevens [108] who considered a population of N units or particles, formed by microorganisms (labeled u) and a chemical produced by them (labeled v), so that $S(N, t) = S_u(N, t) + S_v(N, t)$ denotes the total number of particles in the system at time t . Let $P_N^k(t)$ describe the position of the k th particle ($k \in S(N, t)$) at time t . In [108] the following empirical processes are considered:

$$t \rightarrow S_{N_u}(t) = \frac{1}{N} \sum_{k \in S_u} \delta_{P_N^k(t)}, \quad t \rightarrow S_{N_v}(t) = \frac{1}{N} \sum_{k \in S_v} \delta_{P_N^k(t)},$$

where δ_x denotes Dirac's delta at $x \in \mathbb{R}^d$, $d \geq 1$. A key assumption in [108] is that the dynamics of each particle depends on the configuration of other particles in a neighborhood around it, and as $N \rightarrow \infty$, the interaction of particles is rescaled in a moderate way (cf. Section 2 in [108]). On introducing smoothed versions of $S_{N_i}(t)$, $i = u, v$, of the form $\widehat{S}_{N_i}(t, x) = (S_{N_i}(t) * W_N * \widehat{W}_N)(x)$ for some probability densities $W_N, \widehat{W}_N$ satisfying a suitable scaling assumption (which provides a precise meaning to the assumption on moderate interaction), a system of stochastic differential equations is written for the particle populations (cf. Section 3 in [108]). Using Ito's formula (cf., for instance, [38, 116]), one eventually obtains for any regular test function f ,

$$\begin{aligned} & \langle S_{N_u}(t), f \rangle \\ &= \frac{1}{N} \sum_{k \in S_u} f(t, P_N^k(t)) \\ &= \langle S_{N_u}(0), f(0, \cdot) \rangle \\ &+ \int_0^t \left\langle S_{N_u}(s), \chi_N(s, \cdot) \nabla \hat{S}_{N_v}(s, \cdot) \nabla f(s, \cdot) + \mu \Delta f(s, \cdot) + \frac{\partial}{\partial s} f(s, \cdot) \right\rangle ds \\ &+ \frac{1}{N} \int_0^t \sum_{k \in S_u} \sqrt{2\mu} \nabla f(s, P_N^k(s)) dW^k(s), \end{aligned} \quad (134)$$

$$\begin{aligned} & \langle S_{N_v}(t), f \rangle \\ &= \langle S_{N_v}(0), f(0, \cdot) \rangle \\ &+ \int_0^t \left\langle S_{N_v}(s), \eta \Delta f(s, \cdot) + \frac{\partial}{\partial s} f(s, \cdot) \right\rangle ds \\ &+ \frac{1}{N} \int_0^t \sum_{k \in S_v} \sqrt{2\eta} \nabla f(s, P_N^k(s)) dW^k(s) + \frac{1}{N} \int_0^t \sum_{k \in S_u} f(s, P_N^k(s)) \beta_N^{k*}(ds) \\ &- \frac{1}{N} \int_0^t \sum_{k \in S_v} f(s, P_N^k(s)) \gamma_N^{k*}(ds). \end{aligned} \quad (135)$$

In equations (134), (135) it is assumed that any particle $k \in S_u(N, t)$ at position $P_N^k(t)$ at time t may produce a particle $k^* \in S_v(N, t)$ with intensity $\beta_N(t, P_N^k(t))$, where $\beta_N(t, x) = \beta(\hat{S}_{N_u}(t, x), \hat{S}_{N_v}(t, x))$. This reflects the assumption that chemoattractant is produced by the chemotactic cells themselves. On the other hand, any particle $k \in S_v(N, t)$ may decay with a coefficient $\gamma_N(t, P_N^k(t)) = \gamma(\hat{S}_{N_u}(t, x), \hat{S}_{N_v}(t, x))$. On its turn, $\chi_N(t, P_N^k(t)) = \chi(\hat{S}_{N_u}(t, x), \hat{S}_{N_v}(t, x))$ is a chemosensitivity term arising from the stochastic equation

$$dP_N^k(t) = \chi_N(t, P_N^k(t)) \nabla \hat{S}_{N_v}(t, P_N^k(t)) dt + \sqrt{2\mu} dW^k(t),$$

where $W^k(\cdot)$ are independent Brownian movements. Finally, $\eta > 0$ and $\mu > 0$, and $\beta_N^{k*}(\sigma)$ and $\gamma_N^{k*}(\sigma)$ are taken to be Poisson-type point processes (cf. Section 3 in [108]).

Under a number of technical assumptions, it was proved in [108] that one can pass to the limit in (134), (135) as $N \rightarrow \infty$, to eventually obtain the following limit equations:

$$\begin{aligned}
 & \langle u(t, \cdot), f(t, \cdot) \rangle \\
 &= \langle u_0(\cdot), f(0, \cdot) \rangle \\
 &+ \int_0^t \langle u(s, \cdot), \chi_\infty(s, \cdot) \nabla v(s, \cdot) \nabla f(s, \cdot) \rangle ds \\
 &+ \int_0^t \left\langle u(s, \cdot), \mu \Delta f(s, \cdot) + \frac{\partial}{\partial s} f(s, \cdot) \right\rangle ds, \\
 & \langle v(t, \cdot), f(t, \cdot) \rangle \\
 &= \langle v_0(\cdot), f(0, \cdot) \rangle \\
 &+ \int_0^t \left\langle v(s, \cdot), \mu \Delta f(s, \cdot) + \frac{\partial}{\partial s} f(s, \cdot) - \gamma_\infty(s, \cdot) f(s, \cdot) \right\rangle ds \\
 &+ \int_0^t \langle u(s, \cdot), \beta_\infty(s, \cdot) f(s, \cdot) \rangle ds,
 \end{aligned}$$

where $\chi_\infty(\tau, x) = \chi(u(x, t), v(x, t))$ and a similar definition is made for $\beta_\infty, \gamma_\infty$. These equations can be considered as a weak form of

$$\begin{aligned}
 \frac{\partial u}{\partial t} &= \mu \Delta u - \nabla(\chi(u, v) \nabla v), \\
 \frac{\partial v}{\partial t} &= \eta \Delta u + \beta(u, v)u - \gamma(u, v)v,
 \end{aligned}$$

which is of Keller–Segel type (cf. (101), (102)).

3.4. Pattern formation in *Dictyostelium discoideum*

Here we shall comment on some aspects of the starvation-induced aggregation in monolayer colonies of the slime mold *Dd*. This is a well-documented phenomenon, for which a wealth of evidence is available. We shall recall below some key features in that process, and some related mathematical models will be remarked upon.

(I) *The dynamics of aggregation: some facts.* In this paragraph we shall follow the recent survey [48] and describe the early stages of aggregation in sequential order. The aggregation process induced by food exhaustion last for about 8 hours in wild type (WT) *Dd* cultures. For aggregation to start, a minimum cell density (about 2.5×10^4 cells/cm²) is required. Laboratory experiments are customarily done at much higher densities, of about $4\text{--}65 \times 10^4$ cells/cm². When the process starts, a few cells, which are supposed to be at a comparatively advanced stage of their own cell cycle, begin to emit pulses of cAMP in

a periodic manner, approximately every 5–6 minutes. Upon reception of the cAMP signal, cells internally produce cAMP, part of which is secreted outside to keep the signaling process going on, thus establishing a feedback loop. In monolayer cultures, cAMP propagation gives rise to target and spiral wave patterns that can be observed by means of darkfield microscopy. Reception of the cAMP signals is followed by migration toward some of the cAMP sources. In doing so, streams of *Dd* cells are formed that converge toward aggregation points. There, a size-regulation process is observed, since aggregates achieve a characteristic size, which depends on the number of cells initially available and the dimensions of the surrounding medium. As condensates are formed, a transition from a bidimensional setting to a three-dimensional one occurs, since cells in the condensates pile up each other to eventually produce solid, full three-dimensional mounds. Formation of the mound is followed by a subsequent stage in the developmental process of the *Dd* colony which will not be considered here.

In the sequence of steps briefly recalled above, a few ones stand out as particularly intriguing. One of those is the onset of circular and spiral waves, and the transition from one type of dynamics to the other. We have mentioned that aggregation starts when a few spots in the colony (each of them possibly containing a reduced number of synchronized cells) begins to emit periodic pulses of cAMP. These are visualized as circular waves, usually termed as targets. Soon after that, darkfield microscopy reveals a coexistence of targets and spiral patterns, and by 5 hours after starvation, spiral wave territories dominate and persist in WT *Dd* colonies. The transition just mentioned is known to depend on a number of parameters. One of them is cell density, since targets dominate for low values, whereas for higher densities the situation is just the opposite. Also, addition of a uniform spray of cAMP has dramatic effects on the nature of the patterns observed, depending on the timing of cAMP supply. For instance, if applied soon after starvation, spirals are temporally suppressed to eventually reappear afterward. However, if the uniform cAMP signal is sprinkled at later times, spirals happen to be suppressed for good, and only targets will remain henceforth (cf. [68,69]; see also [48] for a review of related results). In the sequel we shall concentrate on describing some of the mathematical approaches that have been proposed to account for various aspects of the signaling features just recalled.

(II) *From targets to spirals: mathematical models.* The aggregation picture just sketched raises a number of questions. For instance, one may wonder how target and spiral patterns are generated, and what is the precise manner in which a transition from the first to the second type of waves occurs. These issues have been addressed by a number of authors, in particular by Cox [67,88,101], Goldbeter [66,74] and Othmer [26,85,112] to mention but a few names. While a satisfactory global model for the overall aggregation process has not been obtained as yet, a good deal of knowledge on some particular steps has been already obtained. For this we refer to the aforementioned references as well as to the review [48] where certain of these results are discussed in some detail.

In mathematical terms, a basic preliminary question consists in identifying those reaction–diffusion systems (and the underlying physical assumptions) that admit traveling wave solutions in the form of expanding circles (targets) or spirals. This has been shown to occur both in oscillatory and excitable systems, on which we shortly remark below. To simplify the presentation, we shall confine ourselves to the case of continuous equations,

although discrete models (for instance, of a cellular automaton type) have proven to be quite useful to gain insight into a number of related biological problems. For this last type of techniques, the reader is referred to the recent monograph [30].

Let us briefly remark on terminology. Following [79], we shall consider oscillatory media as a continuous limit of a large population of self-oscillating elements, with (weak) interactions between neighbors due to diffusion. On its turn, excitable media are formed by elements, any of which is able to return to an initial rest state after undergoing a burst of activity triggered by a sufficiently large perturbation, which may be originated by diffusional flow from neighboring elements in the medium. A characteristic of excitable media is that they allow for propagation of pulses, a type of traveling wave which connects the same equilibrium value ahead and behind the wave. Oscillatory and excitable regimes may sequentially develop and coexist in biological systems. We next remark on different aspects of wave propagation in such situations.

A typical model of oscillatory media is provided by the so-called λ - ω systems [52,63] which are of the form

$$\frac{\partial u}{\partial t} = D_1 \Delta u + \lambda(A)u - \omega(A)v, \quad (136)$$

$$\frac{\partial v}{\partial t} = D_2 \Delta v + \omega(A)u + \lambda(A)v, \quad (137)$$

where $D_1, D_2 > 0$, and λ, ω are given functions of $A = (u^2 + v^2)^{1/2}$. On these, conditions are imposed so that the reduced kinetic system (obtained by setting $D_1 = D_2 = 0$) should have a stable limit cycle with amplitude α and frequency $\omega(\alpha)$. When $D_1 = D_2 = D > 0$ in (136), (137) that system may be written in a more compact manner by setting

$$w = u + iv, \quad (138)$$

which yields

$$\frac{\partial w}{\partial t} = (\lambda + i\omega)w + D\Delta w. \quad (139)$$

For instance, in the case considered in [64], $\lambda(a) = \varepsilon - aA^2$, $\omega(a) = c - bA^2$, so that (139) reads

$$\frac{\partial w}{\partial t} = (\varepsilon + ic)w - (a + ib)|w|^2w + D\Delta w,$$

which can be thought of as a particular type of a Ginzburg–Landau equation (cf. [63] for details on the derivation of that type of models). It is natural to look for solutions of (139) in the form

$$w = Ae^{i\phi}, \quad (140)$$

where A is an amplitude variable and ϕ its corresponding phase. From (139) and (140), it readily follows that A and ϕ should satisfy

$$\frac{\partial A}{\partial t} = A\lambda(A) - DA|\nabla\phi|^2 + D\Delta A, \quad (141)$$

$$\frac{\partial \phi}{\partial t} = \omega(A) + \frac{2D}{A}(\nabla A \cdot \nabla \phi) + D\Delta \phi. \quad (142)$$

At this juncture, it is worth observing that a large class of reaction–diffusion equations can be approximated, in some asymptotic limit, by means of λ – ω systems. For instance, let us follow [43] and consider the equations

$$\frac{\partial A_1}{\partial t} = F_1(\mu, A_1, A_2) + \nabla(D_1(\mu, A_1, A_2)\nabla A_1), \quad (143)$$

$$\frac{\partial A_2}{\partial t} = F_2(\mu, A_1, A_2) + \nabla(D_2(\mu, A_1, A_2)\nabla A_2), \quad (144)$$

where μ is a (nondimensional) parameter such that at some value $\mu = \mu_0$ the reduced kinetic equations (obtained by setting $D_1 = D_2 = 0$ above) undergoes a bifurcation from a stable state (A_1^0, A_2^0) to a stable limit cycle: in mathematical terms, a Hopf bifurcation is said to occur. Arguing as in [43], Appendix A, one then assumes $0 < \mu - \mu_0 \ll 1$, and look for solutions of the form

$$A_i \sim A_i^0 + (\mu - \mu_0)^{1/2} A(T, \tilde{x}) a_i \cos(\omega t + \gamma_i + \phi(T, \tilde{x})), \quad i = 1, 2,$$

where $\tilde{x} = (\tilde{x}_1, \tilde{x}_2) = (\mu - \mu_0)^{1/2}(x_1, x_2)$, a_i and γ_i are suitable constants, and $T = (\mu - \mu_0)t$. Then the amplitude ϕ and phase A are shown to evolve according to

$$\begin{pmatrix} \frac{\partial A}{\partial T} \\ A \frac{\partial \phi}{\partial T} \end{pmatrix} = \begin{pmatrix} \cos z & -\sin z \\ \sin z & \cos z \end{pmatrix} \begin{pmatrix} \Delta A - A|\nabla\phi|^2 \\ A\Delta\phi + 2\nabla A \cdot \nabla\phi \end{pmatrix} + \begin{pmatrix} A(1 - A^2) \\ qA^3 \end{pmatrix}, \quad (145)$$

where q and z are certain constants determined from the original system (143), (144). In particular, when $D_1 = D_2$, then $z = 0$, and a λ – ω system is obtained, namely

$$\frac{\partial A}{\partial T} = \Delta A - A|\nabla\phi|^2 + A(1 - A^2), \quad (146)$$

$$A \frac{\partial \phi}{\partial T} = A\Delta\phi + 2\nabla A \cdot \nabla\phi + qA^3. \quad (147)$$

We next discuss on spiral patterns. An m -armed ($m \geq 1$) spiral wave of (139) is defined as a solution of the form (140) (if any), such that

$$A = A(r), \quad \phi = \Omega t + m\theta + \psi(r), \quad (148)$$

where (r, θ) are polar coordinates in $\mathbb{R}^2$. A quick check then reveals that A and ψ should then satisfy

$$D\left(A'' + \frac{A'}{r}\right) + A\left(\lambda(A) - D(\psi')^2 - \frac{Dm^2}{r^2}\right) = 0, \quad (149)$$

$$D\left(\psi'' + \left(\frac{1}{r} + \frac{2A'}{A}\right)\psi'\right) = \Omega - \omega(A), \quad (150)$$

a system which is to be supplemented with boundary conditions

$$A(0) = \psi'(0) = 0, \quad A(r) \rightarrow A(\infty) \quad \text{as } r \rightarrow \infty. \quad (151)$$

From (149)–(151) it follows at once that

$$\psi'(\infty) = \left(\frac{\lambda(A_\infty)}{D}\right)^{1/2}, \quad \Omega = \omega(A_\infty).$$

A brief account of early existence results for (149)–(151) can be found in [46]. A rather general existence result has been obtained in [102] (cf. also [35]) that will be described next. Consider the system

$$\frac{\partial u}{\partial t} = D\Delta u + f(u, \mu) \quad \text{for } x \in \mathbb{R}^2, u = (u_1, u_2). \quad (152)$$

Assume that $f(0, \mu) = 0$ for $0 < |\mu| \ll 1$ and that the linearization $\frac{\partial f}{\partial u}(0, 0)$ has a pair of purely imaginary eigenvalues $\pm i\omega_H$, so that the corresponding Hopf bifurcation in the purely kinetic case ($D = 0$) can be written, in suitable variables, in the form

$$\dot{z} = \lambda(\mu)z + \beta z|z|^2 + O(|z|^5) \quad \text{with } \operatorname{Re} \beta < 0. \quad (153)$$

The linearization of (152) around $u = 0$ reads

$$\frac{\partial u}{\partial t} = D\Delta u + \frac{\partial f}{\partial u}(0, 0)u,$$

which after taking Fourier transform yields the dispersion relation

$$d(\lambda, ik) \equiv \det\left(-Dk^2 + \frac{\partial f}{\partial u}(0, 0) - \lambda\right) = 0 \quad \text{for } k \in \mathbb{R}. \quad (154)$$

Then for wavenumbers $k \in \mathbb{R}$ near zero, the eigenvalue $\lambda = i\omega_H$ continues to a spectral curve $\lambda(ik, 0)$ such that $\lambda(0, 0) = i\omega_H$ and

$$\lambda(ik, 0) = i\omega_H + \alpha k^2 + O(k^4). \quad (155)$$

The following result has been proved in [102]. Assume that (i) $\lambda(0, 0) = i\omega_H$ is a simple zero of (154) and the only purely imaginary solution of that equation for any

real k , (ii) $\frac{\partial}{\partial \mu} \operatorname{Re} \lambda(0, 0) > 0$ and for α, β given respectively in (155) and (153) we have $|\arg(\beta/\alpha)| < \delta$ for some $\delta > 0$ sufficiently small. Then for $\mu > 0$ small enough, there exists an Archimedean spiral wave of (152).

By this last we mean a bounded rotating wave solution $q_*(r, \theta - \omega_* t)$ of (152), with some nonzero rotation frequency ω_* , which converges to plane wavetrains in the farfield, that is,

$$\left| q_*(r, \theta) - q_\infty \left(r - \left(\frac{\theta}{k_\infty} \right) \right) \right| \rightarrow 0$$

as $r \rightarrow \infty$ uniformly for $\theta \in [0, 2\pi]$, for some k_∞ .

We now turn our attention to target patterns. These can be roughly described as a wave train of concentric circles propagating from a center, which is often termed as a pacemaker. Following [42], we shall look for targets in reaction–diffusion systems of the type

$$\frac{\partial A}{\partial t} = F(A) + \varepsilon D \Delta A + \varepsilon g(x, A), \quad (156)$$

where $A = (A_1, A_2)$, $D > 0$, $0 < \varepsilon \ll 1$ and $g(A, x)$ is a bounded function of its arguments. As to the kinetic term $F(A)$, we shall assume that the autonomous ODE system

$$\dot{A} = F(A), \quad (157)$$

has a stable time-periodic solution $B(t) = B(t + P)$ for some $P > 0$. We now introduce a slow-time scale $T = \varepsilon t$, and look for solutions of (156) of the form

$$A(\varepsilon, t, x) = A^0(t, T, x) + \varepsilon A^1(t, T, x) + \varepsilon^2 A^2(t, T, x) + \dots$$

requiring $A^1, A^2, \dots$ to be bounded in time. Substituting this expansion into (156) gives

$$\frac{\partial A^0}{\partial t} = F(A^0), \quad (158)$$

$$\frac{\partial A^1}{\partial t} - \frac{\partial F}{\partial A}(A^0) A^1 = -\frac{\partial A^0}{\partial T} + D \Delta A^0 + g(x, A^0). \quad (159)$$

Solving (158) yields

$$A^0 = B(t + \psi(T, x)), \quad (160)$$

where $\psi(T, x)$ is a phase variable which remains undetermined at this stage. As a matter of fact, plugging (160) into (159), and looking then for bounded solutions in the resulting equation (which requires imposing a suitable orthogonality condition there), one eventually obtains that ψ satisfies

$$\frac{\partial \psi}{\partial T} = D_1 (\Delta \psi + \Gamma |\nabla \psi|^2) + \alpha(x), \quad (161)$$

where

$$\begin{aligned}
 D_1 &= \frac{1}{P} \int_0^P z^T(s) D B'(s) ds, \\
 \Gamma &= \frac{1}{P} \int_0^P z^T(s) D B''(s) ds, \\
 \alpha(x) &= \frac{1}{P} \int_0^P z^T(s) g(x, B(s)) ds,
 \end{aligned} \tag{162}$$

and the vector z^T is periodic with period P and such that

$$\begin{aligned}
 \frac{\partial z^T}{\partial t} + z^T F_A(B(t + \psi)) &= 0, \quad z^T(t + \psi) B'(t + \psi) = 1, \\
 \int_0^P z^T G ds &= 0,
 \end{aligned}$$

where

$$G \equiv G(\psi, x, t) = -B' \frac{\partial \psi}{\partial T} + D(B' \Delta \psi + B'' |\nabla \psi|^2) + g(x, B).$$

Summing up, we have obtained a solution of (156) in the form

$$A(x, t) = B(t + \psi(T, x)) + O(\varepsilon), \tag{163}$$

where ψ solves (161). This representation is consistent with our assumption of a distributed medium consisting in a large population of individual oscillators, weakly coupled by diffusion, which produces a phase shift between different points, whose time evolution is described by (161).

It remains to be seen if (163) provides target patterns for (156). A remarkable fact is that this may be the case only if $g(x, A) \neq 0$, so that inhomogeneities in the medium are crucial for the onset of such type of waves in systems as (156). To check this statement, assume on the contrary that $g(x, A) = 0$. Then, by (162), $\alpha(x) = 0$ and setting $Z = e^{\Gamma \psi}$, (161) reduces to

$$\frac{\partial Z}{\partial T} = D \Delta Z.$$

It then turns out that the initial value problem for (161) can be explicitly solved, and there holds

$$\psi(T, x) = \frac{1}{\Gamma} \log \left((4\pi D_1 T)^{-1} \int_{\mathbb{R}^2} \exp \left(\frac{\Gamma \psi(0, y) - |x - y|^2}{4D_1 T} \right) dy \right).$$

From this formula we see that, if $\psi(0, x)$ is bounded, then $\psi(T, x)$ converges to a constant as $T \rightarrow \infty$, so that asymptotically the medium oscillates with uniform phase shift. When inhomogeneities are present, however, an asymptotic analysis detailed in [42] shows the existence of target patterns for (156), for which the propagation parameters are estimated.

As a next step, we now remark on the coexistence of excitable and oscillatory regimes. This we shall do by following the arguments by Hagan and Cohen in [44]. In that work, a dynamical model for regulation of cAMP in *Dictyostelium discoideum* was proposed in terms of a number of variables: external (respectively internal) cAMP, cAMPe (respectively cAMPi), a cAMP inhibitor and a lump variable accounting for intracellular stored reserves. After performing a suitable asymptotic analysis, made possible by the separation of scales in the model that followed from the consideration of some small parameters, one is essentially led to analyzing the following system:

$$\dot{A} = f(A, C), \quad (164)$$

$$\dot{C} = k(A)S - h_1(C) \equiv g(A, C, S), \quad (165)$$

$$\dot{S} = \varepsilon(h_2(c) - k(A)S), \quad (166)$$

where $0 < \varepsilon \ll 1$, and f, k, h_1, h_2 are bounded functions whose qualitative behavior is depicted in Figure 7. As a matter of fact, small diffusivity effects are considered in [44]

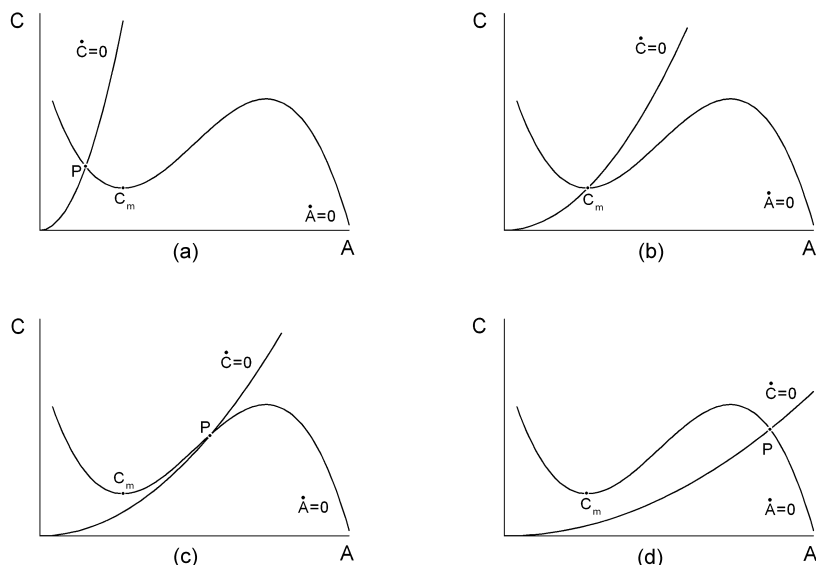


Fig. 7. Sequence of propagation regimes in system (164)–(167): (a) As P approaches C_m , excitability shows in, and cAMP pulses will propagate if triggered by a sufficiently large stimulus; (b) and (c) A Hopf bifurcation signals the entrance in the oscillatory regime. Region (d) corresponds to a stable equilibrium with high rates of synthesis and leakage of A . (Adapted from [44].)

(and should therefore be added to (164)–(166)), but these may be omitted in the forthcoming discussion. For completeness, we point out that variables A , C and S represent scaled versions of the intracellular cAMP, inhibitor and stored reserve concentrations respectively.

In (164)–(166), S can be thought of as a slowly varying parameter in the two-dimensional system consisting of (164) and (165), and its effect results in changes in the nullcline $g(A, C, S) = 0$ in (165). In this way, as S decreases the corresponding phase portrait goes from stages (a)–(d) in Figure 7.

At this juncture, the reader may wonder if one could possibly derive a multiscale model that should be able to account for individual behavior (including periodic firing at some pacemakers) on the one hand, and at the same time reproduce the transition and coexistence of macroscopic patterns as targets and spirals at the other extreme. At some stage, any such model is likely to involve an effective medium (or homogenization) approach. While no such model seems to be available as yet (cf. in this sense the discussion in [48]), it might be of some interest to shortly remark on the point of view recently developed in [101]. In that work, experiments on *Dd* mutant and wild-type (WT) strains are reported. A remarkable fact is that both mutant and WT cultures display optical density oscillations, although at a different pace in each case. In particular, the ability of mutant strains to produce self-organizing spiral patterns is seriously diminished, although the oscillation kinetics in all strains seem to be quite similar. A quantitative discussion on the spiral patterning is provided in [101], supplementary material. In doing so, the authors made use of a model system given by

$$\frac{\partial C}{\partial t} = D\Delta C - \Gamma C + H(C - C_T)C_r, \quad (167)$$

$$C_T = \left(C_{\max} - \frac{At}{t + T} \right) (1 - E), \quad (168)$$

$$\frac{\partial E}{\partial t} = \eta + \beta C. \quad (169)$$

Here $C = C(x, y, t)$ stands for the concentration of *Dd* amoebae, which are assumed to be in one of three possible states: excitable, excited and refractory. When excitable cells are subject to a cAMP concentration exceeding a value C_T , they become excited and release a cAMP pulse C_r . After that, cells enter into a refractory state where no cAMP is secreted. In that period, C_T decreases from a value $C_{\max}$ to a value $C_{\min} < C_{\max}$ after a time $t = \tau > 0$ when they enter again into the excitable state. Excitability is represented in (167)–(169) by variable E ; D , T , A , η , β are various positive parameters, and $H(s)$ is Heaviside function: $H(s) = 1$ if $s > 0$, $H(s) = 0$ otherwise. In particular, it was observed in [101] that when $\beta < 10^{-3}$ (low excitability) multiple firing centers appear, whereas for $\beta > 10^{-2}$ (high excitability) spiral waves are observed to persist. An interesting question would be to derive the evolution in time of β from data corresponding to interaction of *Dd* cells at the microscopic level.

Acknowledgements

This work has been supported by European Contract MRTN-CT-2004-503661 and by Acción Especial AE5/06-14364 from Universidad Complutense.

References

- [1] T. Alarcón and K.M. Page, *Stochastic models of receptor oligomerization by bivalent ligand*, J. Roc. Soc. Interface (2006), published online, DOI: 10.1098/rsif.2006.0116.
- [2] T. Alarcón and K.M. Page, *Stochastic models of the VGEF receptor: Analysis and implications on anti-VGEF cancer therapy*, J. Roc. Soc. Interface (2006), submitted.
- [3] R. Albert, Y.W. Chiu and H.G. Othmer, *Dynamic receptor team formation can explain the high signal transduction gain in Escherichia coli*, Biophysical J. **86** (2004), 2650–2659.
- [4] D. Ambrosi, A. Gamba, E. Giraud, G. Serini, L. Preziosi and F. Bussolino, *Burgers dynamics governs the early stages of vascular network assembly*, EMBO J. **22** (2003), 1771–1779.
- [5] H. Baier and F. Bonhoeffer, *Axon guidance by gradients of a target-derived component*, Science **255** (1992), 472–475.
- [6] N. Barkai and S. Leibler, *Robustness in simple biochemical networks*, Nature **387** (1997), 913–917.
- [7] N. Bellomo, A. Bellouquid and M. Delitala, *Mathematical topics on the modeling of multicellular systems in the competition between tumor and immune cells*, Math. Models Methods Appl. Sci. **14** (2004), 1683–1733.
- [8] N. Bellomo, A. Bellouquid and M.A. Herrero, *From microscopic to macroscopic description of multicellular systems and biological growing tissues*, Math. Cont. Mech., to appear.
- [9] H.C. Berg, *Random Walks in Biology*, Princeton Univ. Press, Princeton, NJ, USA (1993).
- [10] H.C. Berg, *E-coli in Motion*, Springer-Verlag, New York, USA (2003).
- [11] H.C. Berg and E.M. Purcell, *Physics of chemoreception*, Biophysical Journal **20** (1977), 193–219.
- [12] P. Biler, *Local and global solvability of some parabolic systems modeling chemotaxis*, Adv. Math. Sci. Appl. **9** (1998), 347–359.
- [13] P. Biler, G. Karch, Ph. Laurençot and T. Nadzieja, *The 8π -problem for radially symmetric solutions of a chemotaxis model in the plane*, Preprint (2006).
- [14] P. Biler, G. Karch, Ph. Laurençot and T. Nadzieja, *The 8π -problem for radially symmetric solutions of a chemotaxis model in a disc*, Preprint (2006).
- [15] J.T. Bonner, *The Cellular Slime Mold*, Princeton Univ. Press, Princeton, NJ, USA (1967).
- [16] J.T. Bonner, *Sixty Years of Biology*, Princeton Univ. Press, Princeton, NJ, USA (1996).
- [17] D. Bray, M.D. Levin and C.J. Morton-Firth, *Receptor clustering as a cellular mechanism to control sensitivity*, Nature **393** (1998), 85–88.
- [18] F.A. Chalub, P. Markowich, B. Perthame and C. Schmeiser, *Kinetic models for chemotaxis and their drift-diffusion limits*, Monatsh. Math. **142** (2004), 123–141.
- [19] S. Chandrasekhar, *Stochastic problems in physics and astronomy*. Rev. Mod. Phys. **15** (1943), 1–91.
- [20] N.B. Charvet, K. Brose, K. Wang, V. Marillat, T. Kidd, C.S. Goodman, M. Tessier-Lavigne, C. Sotelo and A. Chedotal, *Slit-2 mediated chemorepulsion and collapse of developing forebrain axons*, Neuron **22** (1999), 463–473.
- [21] S. Childress, *Chemotactic collapse in two dimensions*, Lecture Notes in Biomathematics (Springer) **55** (1984), 61–66.
- [22] S. Childress and J.K. Percus, *Nonlinear aspects of chemotaxis*, Math. Biosci. **56** (1981), 217–237.
- [23] R.L. Chisholm and R.A. Firtel, *Insights into morphogenesis from a simple developmental system*, Nature Rev. Mol. Cell Biol. **5** (2004), 531–541.
- [24] L. Corrias, B. Perthame and H. Zaag, *Global solutions of some chemotaxis and angiogenesis systems in high space dimensions*, Milan J. Math. **72** (2004), 1–28.
- [25] J. Crank, *The Mathematics of Diffusion*, Oxford Univ. Press, Oxford, UK (2003).
- [26] J.C. Dallon and H.G. Othmer, *A continuum analysis of the chemotactic signal seen by Dictyostelium discoideum*, J. Theoret. Biol. **194** (1998), 461–483.

- [27] F. de Castro, *Chemotropic molecules: Guides for axonal pathfinding and cell migration during CNS development*, News Physiol. Sci. **18** (2003), 130–136.
- [28] C. De Lisi, F. Marchetti and G. Del Grosso, *A theory for measurement error and its implications for spatial and temporal gradient sensing during chemotaxis*, Cell Biophys. **4** (1982), 211–229.
- [29] C. De Lisi, and F. Marchetti, *A theory for measurement error and its implications for spatial and temporal gradient sensing during chemotaxis, II: The effects of non-equilibrated ligand binding*, Cell Biophys. **5** (1983), 237–253.
- [30] A. Deutsch and S. Dormann, *Cellular Automaton Modeling of Biological Pattern Formation*, Birkhäuser, Boston, MA, USA (2005).
- [31] M. Eisenbach, ed., *Chemotaxis*, Imperial College Press, London, UK (2004).
- [32] R. Erban and H.G. Othmer, *From individual to collective behaviour in bacterial chemotaxis*, SIAM J. Appl. Math. **65** (2004), 361–391.
- [33] R. Erban and H.G. Othmer, *From signal transduction to spatial pattern formation in E. coli: A paradigm for multi-scale modeling in biology*, Multiscale Model. Simul. **3** (2) (2005), 362–394.
- [34] R. Escalante and J.J. Vicente, *Dictyostelium discoideum: A model system for differentiation and patterning*, Int. J. Devel. Biol. **44** (2000), 819–835.
- [35] B. Fiedler and A. Scheel, *Spatio-temporal dynamics of reaction–diffusion patterns*, Trends in Nonlinear Analysis, Springer-Verlag, Berlin (2002), 21–150.
- [36] S.K. Friedlander, *Smoke, Dust and Haze: Fundamentals of Aerosol Dynamics*, Oxford Univ. Press, New York, USA (2000).
- [37] A. Gamba, D. Ambrosi, A. Coniglio, A. de Candia, S. DiTalia, E. Giraudo, G. Serini, L. Preziosi and F. Bussolino, *Percolation, morphogenesis and Burgers dynamics in blood vessels formation*, Phys. Rev. Lett. **90** (2003), 11810–11814.
- [38] C.W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer-Verlag (1997).
- [39] A. Gierer and H. Meinhardt, *A theory of biological pattern formation*, Kybernetik **12** (1972), 30–39.
- [40] G.J. Goodhill and J.S. Urbach, *Theoretical analysis of gradient detection by growth cones*, J. Neurobiol. **41** (1999), 230–241.
- [41] R. Gordon and L. Belousov, *From observations to paradigms: The importance of theories and models. An interview with Hans Meinhardt*, Int. J. Devel. Biol. **50** (2006), 103–111.
- [42] P.S. Hagan, *Target patterns in reaction–diffusion systems*, Adv. in Appl. Math. **42** (1981), 762–786.
- [43] P.S. Hagan, *Spiral waves in reaction–diffusion equations*, SIAM J. Appl. Math. **42** (1982), 762–786.
- [44] P.S. Hagan and M.S. Cohen, *Diffusion-induced morphogenesis in the development of Dictyostelium*, J. Theoret. Biol. **93** (1981), 881–908.
- [45] R. Heinrich, B.G. Neel and T.A. Rapoport, *Mathematical models of protein kinase signal transduction*, Molecular Cell **9** (2002), 957–970.
- [46] M.A. Herrero, *Reaction–diffusion systems: A mathematical biology approach*, Cancer Modelling and Simulations, L. Preziosi, ed., Chapman & Hall (2003), 367–420.
- [47] M.A. Herrero, E. Medina and J.J.L. Velázquez, *Finite-time aggregation into a single point in a reaction–diffusion system*, Nonlinearity **10** (1997), 1754–1793.
- [48] M.A. Herrero and L. Sastre, *Models of aggregation in Dictyostelium discoideum: On the track of spiral waves*, Networks and Heterogeneous Media **1** (2) (2006), 241–258.
- [49] M.A. Herrero and J.J.L. Velázquez, *Chemotactic collapse for the Keller–Segel model*, J. Math. Biol. **35** (1996), 177–196.
- [50] T. Hillen and H.G. Othmer, *The diffusion limit of transport equations derived from velocity-jump processes*, SIAM J. Appl. Math. **61** (3) (2000), 751–775.
- [51] D. Horstmann, *From 1970 until present: The Keller–Segel model in chemotaxis and its consequences I*, Jahresber. Deutsch. Math.-Verein. **105** (3) (2003), 103–165.
- [52] L.N. Howard and N. Koppel, *Slowly varying waves and shock structures in reaction–diffusion equations*, Stud. Appl. Math. **56** (1977), 95–145.
- [53] P.A. Iglesias and A. Levchenko, *Modelling the cell's guidance system*, Science STKE, available at <http://stke.sciencemag.org/cgi/content/full/sigtransj2002/148/rel2> (2002).
- [54] W. Jäger and S. Luckhaus, *On explosions of solutions to a system of partial differential equations modeling chemotaxis*, Trans. Amer. Math. Soc. **239** (1992), 817–824.

- [55] F. John, *Partial Differential Equations*, Appl. Math. Sci., Vol. 1, Springer-Verlag, New York, USA (1980).
- [56] E.F. Keller, *Making Sense of Life: Explaining Biological Development with Models, Metaphors and Machines*, Harvard Univ. Press, Cambridge, MA, USA (2002).
- [57] E.F. Keller and L.A. Segel, *Initiation of slime mold aggregation viewed as an instability*, J. Theoret. Biol. **26** (1970), 399–415.
- [58] S.H. Kim, W. Wang and K.K. Kim, *Dynamic and clustering model of bacterial chemotaxis receptors: Structural basis for signaling and high sensitivity*, Proc. Natl. Acad. Sci. USA **99** (18) (2002), 11611–11615.
- [59] T.M. Konijn, J.G.C. van de Meere, J.T. Bonner and D.S. Barkley, *The acrasin activity is adenosin – 3',5'-cyclic phosphate*, Proc. Natl. Acad. Sci. USA **58** (1967), 1152–1154.
- [60] R. Kowalczyk, *Preventing blow-up in a chemotaxis model*, J. Math. Anal. Appl. **305** (2005), 566–580.
- [61] R. Kowalczyk, A. Gamba and L. Preziosi, *On the stability of homogeneous solutions to some aggregation models*, Discrete Contin. Dyn. Syst. (4) **13** (2004), 204–220.
- [62] J. Krishnan, P.A. Iglesias and L. Ma, *Spatial sensing of chemotactic gradients: A reaction–diffusion model*, Proc. 2nd Internat. Conf. System Biology, Pasadena, CA (2001), 148–157.
- [63] Y. Kuramoto, *Chemical Oscillations, Waves and Turbulence*, Springer-Verlag, Berlin (1984).
- [64] Y. Kuramoto and S. Koga, *Turbulized rotating chemical waves*, Progr. Theoret. Phys. **66** (1981), 1081–1085.
- [65] D.A. Lauffenburger and J.J. Linderman, *Receptors: Models for Binding, Trafficking and Signalling*, Oxford Univ. Press, Oxford, UK (1993).
- [66] J. Lauzeral, J. Halloy and A. Goldbeter, *Desynchronization of cells on the developmental path triggers the formation of spiral waves of cAMP during Dictyostelium aggregation*, Proc. Natl. Acad. Sci. USA **94** (1997), 9153–9158.
- [67] K.J. Lee, E.C. Cox and R.E. Goldstein, *Competing patterns of signalling activity in Dictyostelium discoideum*, Phys. Rev. Lett. **76** (7) (1996), 1174–1177.
- [68] K.J. Lee, R.E. Goldstein and E.C. Cox, *Resetting wave forms in Dictyostelium territories*, Phys. Rev. Lett. **87** (6) (2001), 0681011.
- [69] K.J. Lee, R.E. Goldstein and E.C. Cox, *cAMP waves in Dictyostelium territories*, Nonlinearity **15** (2002), C1–C5.
- [70] A. Levchenko and P.A. Iglesias, *Models of eukaryotic gradient sensing: Application to chemotaxis of amoebae and neutrophils*, Biophys. J. **82** (2002), 50–63.
- [71] M.D. Levin, T.S. Shimizu and D. Bray, *Binding and diffusion of CheR molecules within a cluster of membrane receptors*, Biophys. J. **82** (2002), 1809–1817.
- [72] S. Luckhaus and Y. Sugiyama, *Asymptotic profile with the optimal convergence rate for a parabolic equation of chemotaxis in super-critical cases*, Preprint (2006).
- [73] M. Maeda, *Regulation of growth and differentiation in Dictyostelium*, Int. Rev. Cytolog. **244** (2005), 287–332.
- [74] J.L. Martiel and A. Goldbeter, *A model based on receptor desensitization for cyclic-AMP signalling in Dictyostelium cells*, Biophys. J. **52** (1987), 807–828.
- [75] J.M. Mato, A. Losada, V. Nanjundiah and T.M. Konijn, *Signal input for a chemotactic response in the cellular slime mold Dictyostelium discoideum*, Proc. Natl. Acad. Sci. USA **72** (1975), 4991–4993.
- [76] H. Meinhardt, *Models of Biological Pattern Formation*, Academic Press, Manchester, UK (1982).
- [77] H. Meinhardt, *Orientation of chemotactic cells and growth cones: Models and mechanisms*, J. Cell Sci. **112** (1999), 2867–2874.
- [78] H. Meinhardt, *Out-of-phase oscillations and traveling waves with unusual properties: The use of three-component systems in biology*, Physica D **199** (2003), 264–277.
- [79] A.S. Mikhailov, *Foundations of Synergetics I*, Springer-Verlag, New York (1994).
- [80] T. Nagai, *Blow-up of radially symmetric solutions to a chemotaxis system*, Adv. Math. Sci. Appl. **5** (1995), 1–21.
- [81] T. Nagai, T. Senba and T. Suzuki, *Chemotaxis collapse in a parabolic system of mathematical biology*, Hiroshima Math. J. **30** (2000), 463–497.
- [82] V. Nanjundiah, *Chemotaxis, signal relaying and aggregation morphology*, J. Theoret. Biol. **42** (1973), 63–105.

- [83] H.G. Othmer, S.R. Dunbar and W. Alt, *Models of dispersal of biological populations*, J. Math. Biol. **26** (1998), 263–298.
- [84] H.G. Othmer and T. Hillen, *The diffusion limit of transport equations, II: Chemotaxis equations*, SIAM J. Appl. Math. **62** (4) (2002), 1222–1250.
- [85] H.G. Othmer and P. Schaap, *Oscillating signaling in the development of Dictyostelium discoideum*, Comments. Theor. Biol. **5** (1998), 175–282.
- [86] H.G. Othmer and A. Stevens, *Aggregation, blow-up and collapse. The ABC's of generalized taxis*, SIAM J. Appl. Math. **57** (1997), 1044–1081.
- [87] K.J. Painter and T. Hillen, *Volume-filling and quorum-sensing in models for chemosensitive movement*, Canad. Appl. Math. Quart. **10** (4) (2004), 501–543.
- [88] E. Palsson and E.C. Cox, *Origin and evolution of circular waves and spirals in Dictyostelium discoideum territories*, Proc. Natl. Acad. Sci. USA **93** (1996), 1151–1155.
- [89] A.S. Perelson, *Receptor clustering on a cell surface II. Theory of receptor cross-linking by ligands bearing two chemically distinct functional groups*, Math. Biosci. **49** (1980), 87–110.
- [90] A.S. Perelson, *Receptor clustering on a cell surface, III. Theory of receptor cross-linking by multivalent ligands: Description of ligand states*, Math. Biosci. **53** (1981), 1–39.
- [91] A.S. Perelson and C. De Lisi, *Receptor clustering on a cell surface I. Theory of receptor cross-linking by ligands bearing two chemically identical functional groups*, Math. Biosci. **48** (1980), 71–110.
- [92] B. Perthame, *PDE models for chemotactic movements: Parabolic, hyperbolic and kinetics*, Appl. Math. **49** (2004), 539–564.
- [93] R.G. Posner, C. Wofsy and B. Goldstein, *The kinetics of bivalent ligand-bivalent receptor aggregation: Ring formation and the breakdown of equivalent site approximation*, Math. Biosci. **126** (1995), 171–190.
- [94] S. Ramón y Cajal, *La retine des vertébrés*, La Cellule **9** (1893), 119–255.
- [95] S. Ramón y Cajal, *Nouvelles observations sur l'évolution des neuroblastes avec quelques remarques sur l'hypothèse neurogénétique de Hensen–Held*, Anat. Anzeiger Bd. XXXII (1908).
- [96] S. Ramón y Cajal, *Estudios sobre la degeneración y regeneración del sistema nervioso*, Vols 1 and 2, Moya, Madrid (1913–1914).
- [97] K.B. Raper, *Dictyostelium discoideum, a new species of slime mold from decaying forest leaves*, J. Agr. Res. **50** (1935), 135–147.
- [98] W.J. Rappel, P.J. Thomas, H. Levine and W.F. Loomis, *Establishing direction during chemotaxis in eukaryotic cells*, Biophys. J. **83** (2002), 1361–1367.
- [99] C. Roisin-Bouffey, W. Jang, D.R. Caprette and R.H. Gomer, *A precise group size in Dictyostelium is generated by a cell-counting factor modulating cell–cell adhesion*, Mol. Cell **6** (2000), 953–959.
- [100] P.G. Saffman and M. Delbrück, *Brownian motion in biological membranes*, Proc. Natl. Acad. Sci. USA **72** (8) (1975), 3111–3113.
- [101] S. Sawal, P.A. Thomson and E.C. Cox, *An autoregulatory circuit for long-range self-organization in Dictyostelium cell populations*, Nature **433** (2005), 323–326.
- [102] A. Scheel, *Bifurcation to spiral waves in reaction–diffusion systems*, SIAM J. Math. Anal. **29** (1998), 1399–1418.
- [103] J.E. Segall, S.M. Block and H.C. Berg, *Temporal comparisons in bacterial chemotaxis*, Proc. Natl. Acad. Sci. USA **83** (1986), 8987–8991.
- [104] T. Serafini, T.E. Kennedy, M.J. Galko, C. Mirzayan, T.M. Jessell and M. Tessier-Lavigne, *The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC–6*, Cell **78** (1994), 409–424.
- [105] C. Sotelo, *The chemotactic hypothesis of Cajal: A century behind*, Progr. Brain Res. **136** (2002), 11–20.
- [106] V. Sourjik and H.C. Berg, *Receptor sensitivity in bacterial chemotaxis*, Proc. Natl. Acad. Sci. USA **99** (2002), 123–127.
- [107] P. Spiro, J.S. Parkinson and H.G. Othmer, *A model of excitation and adaptation in bacterial chemotaxis*, Proc. Natl. Acad. Sci. USA **94** (1997), 7263–7268.
- [108] A. Stevens, *Derivation of chemotaxis equations as limit dynamics of moderately interacting stochastic many particle systems*, SIAM J. Appl. Math. **61** (1) (2000), 183–212.
- [109] Y. Sugiyama, *Global existence in sub-critical cases and finite time blow-up in supercritical cases to degenerate Keller–Segel systems*, J. Differential Equations (2006), to appear.

- [110] Y. Sugiyama and H. Kunii, *Global existence and decay properties for a degenerate Keller–Segel model with a power factor in drift term*, J. Differential Equations (2006), to appear.
- [111] T. Suzuki, *Free-energy and self-interacting particles*, Progr. Nonlinear Differential Equations, Vol. 62, Birkhäuser, Boston, USA (2005).
- [112] Y. Tang and H.G. Othmer, *Excitation, oscillations and wave propagation in a G-protein based model of signal transduction in Dictyostelium discoideum*, Philos. Trans. Roy. Soc. London Ser. B **349** (1995), 179–195.
- [113] M. Tessier-Lavigne, M. Placzek, A.G. Lumsden, J. Dodd and T.M. Jessell, *Chemotropic guidance of developing axons in the mammalian central nervous system*, Nature **336** (1988), 75–778.
- [114] A.M. Turing, *The chemical basis of morphogenesis*, Philos. Trans. Roy. Soc. London **237** (1952), 37–72.
- [115] P.J.M. van Haastert and P.N. Devreotes, *Chemotaxis: Signalling the way forward*, Nature Rev. Mol. Cell Biol. **5** (2004), 626–634.
- [116] A. van Oojen, *Modeling Neural Development*, MIT Press, Cambridge, MA, USA (2003).
- [117] J.J.L. Velázquez, *Point dynamics in a singular limit of the Keller–Segel model, I. Motion of the concentration regions*, SIAM J. Appl. Math. **64** (4) (2004), 1198–1223.
- [118] J.J.L. Velázquez, *Point dynamics in a singular limit of the Keller–Segel model, II. Formation of the concentration regions*, SIAM J. Appl. Math. **64** (4) (2004), 1224–1248.
- [119] D. Wzrosek, *Global attractor for a chemotaxis model with prevention of overcrowding*, Preprint (2004).
- [120] T.M. Yi, Y. Huang, M.L. Simon and J. Doyle, *Robust perfect adaptation in bacterial chemotaxis through integral feedback control*, Proc. Natl. Acad. Sci. USA **97** (9) (2000), 4649–4653.
- [121] S.H. Zigmond, *Ability of polymorphonuclear leukocytes to orient in gradients of chemotactic factors*, J. Cell. Biol. **75** (1977), 606–616.
- [122] R. Zwanzig, *Diffusion-controlled ligand binding to spheres partially covered by receptors: An effective medium treatment*, Proc. Natl. Acad. Sci. USA **87** (1990), 5856–5857.

CHAPTER 3

Examples of Singular Limits in Hydrodynamics

Nader Masmoudi

Courant Institute, New York University, 251 Mercer Street, New York, NY 10012-1185, USA

E-mail: masmoudi@cims.nyu.edu

Contents

1. Introduction	197
1.1. Dimensionless parameters	197
1.2. Mathematical problems	198
1.3. The compressible Navier–Stokes system	198
1.4. Dimensionless parameters	199
2. The inviscid limit	201
2.1. The whole space case	202
2.2. The case of the Dirichlet boundary condition	205
2.3. Weak limit	212
3. Compressible–incompressible limit	215
3.1. Formal limit	216
3.2. The case of strong solutions	217
3.3. The case of global weak solutions	223
3.4. Study of the limit $\gamma \rightarrow \infty$	233
3.5. The nonisentropic case	236
4. Study of rotating fluids at high frequency	237
4.1. The periodic case	238
4.2. Ekman boundary layers in $\Omega = \mathbb{T}^2 \times]0, h[$	238
4.3. The case of other geometries	244
4.4. Other related problems	244
5. Hydrodynamic limit of the Boltzmann equation	245
5.1. Scalings and formal asymptotics	245
5.2. The convergence toward the incompressible Navier–Stokes–Fourier system	251
5.3. The convergence toward the Stokes system	254
5.4. The case of a bounded domain	256
5.5. Convergence toward the Euler system	260
6. Some homogenization problems	261
6.1. Darcy law	264
6.2. Homogenization of a compressible model	264
6.3. Homogenization of the Euler system	266

HANDBOOK OF DIFFERENTIAL EQUATIONS

Evolutionary Equations, volume 3

Edited by C.M. Dafermos and E. Feireisl

© 2007 Elsevier B.V. All rights reserved

7. Conclusion 268

 7.1. Other limits 268

 7.2. Concluding remarks 268

Acknowledgement 269

References 269

Abstract

This chapter is devoted to the study of some asymptotic problems in hydrodynamics. In particular, we will review results about the inviscid limit, the compressible–incompressible limit, the study of rotating fluids at high frequency, the hydrodynamic limit of the Boltzmann equation as well as some homogenization problems in fluid mechanics.

1. Introduction

Any physical system can be described by a system of equations which governs the evolution of the different physical quantities such as the density, the velocity, the temperature The unknowns usually involve several physical units such as (m, kg, s . . .). Introducing some length scale, time scale, velocity scale . . . , the system of equations can always be written in a dimensionless form. This dimensionless form contains some ratios between the different scales such as the Reynolds number, the Mach number or the ratio between two length scales. Indeed, the system may have different length scales. For instance, it may have a vertical length scale and a horizontal one.

1.1. Dimensionless parameters

Writing the system in its dimensionless form allows us to compare the relative influence of the several terms appearing in the equations. Moreover, it allows us to compare different systems. For instance two incompressible flows which have the same Reynolds number have very similar properties, even if the length scales, the velocity scales and viscosities are very different. The only important factor of comparison is the ratio $Re = \underline{U} \underline{L} / \nu_0$ where $\underline{U}$ is the velocity scale, $\underline{L}$ is the length scale, and ν_0 is the kinematic viscosity.

In hydrodynamics, asymptotic problems arise when a dimensionless parameter ε goes to zero in a dimensionless system of equations describing the motion of some fluid. Physically, this allows a better knowledge of the system in this limit regime by describing (usually by a simpler system) the prevailing phenomenon when this parameter is small. Indeed, this small parameter, usually describes a physical reality. For instance, a slightly compressible flow is characterized by a low Mach number, whereas a slightly viscous flow is characterized by a high Reynolds number. Notice, here, that we used the terminology slightly compressible flow or slightly viscous flow instead of fluid. Indeed, this is a property of the flow rather than the fluid itself. However, we will often use the terminology slightly compressible fluid or slightly viscous fluid to mean the properties of the flow.

Let us notice that if the viscosity goes to zero, then the Reynolds number goes to infinity. But this is not the only way of getting a big Reynolds number. For instance, if $\underline{L}$ or $\underline{U}$ increase then the Reynolds number also increases and we get the same properties as when the viscosity goes to zero. This is of course very important from a physical point of view since it is much easier to change $\underline{L}$ or $\underline{U}$ in a physical experiment than to change the viscosity. This shows the importance of the dimensionless parameters. So, when we speak about the inviscid limit, this should be understood as the limit when the Reynolds number goes to infinity.

Moreover, in many cases, we have different small parameters (we can be in presence of a slightly compressible and slightly viscous fluid in the same time). Depending on the way these small parameters go to zero, we can recover different systems at the limit. For instance, if $\varepsilon, \delta, \nu, \eta \ll 1$, the limit system can depend on the magnitude of the ratio of ε/δ or ε/ν This again shows the importance of having dimensionless quantities which can be compared.

The study of these asymptotic problems allows us to get simpler models at the limit, due to the fact that we usually have fewer variables or (and) fewer unknowns. This simplifies the numerical simulations. In fact, instead of solving the initial system, we can solve the limit system and then add a corrector.

1.2. Mathematical problems

Many mathematical problems are encountered when we try to justify the passage to the limit, which are mainly due to the change of the type of the equations the presence of many spatial and temporal scales, the presence of boundary layers (we can no longer impose the same boundary conditions for the initial system and the limit one), the presence of oscillations in time at high frequency

Usually, we say that we have a singular limit if there is a change of the type of the equation. For instance in the inviscid limit (Reynolds number going to infinity), we go from a parabolic equation to a hyperbolic equation. However, this terminology seems a little bit restrictive since, we can see from the examples that it is not usually easy to give a type to each system of equations. Moreover, we can say that we have a singular limit if we have a reduction of the number of variables or unknowns due to a more restrained dynamics. Different type of questions can be asked:

1. What do the solutions of the initial system (S_ε) converge to? Is the convergence strong or weak?
2. In the case of weak convergence, can we give a more detailed description of the sequences of solutions? Can we describe the time oscillations for instance?
3. Can we use some properties of the limit system to deduce properties for the initial system when the parameter is small.

In this chapter we will try to answer some of these questions by studying some examples of singular limits in hydrodynamics. In the next subsection we recall the physical equation of fluid dynamics and introduce the several dimensionless parameters.

1.3. The compressible Navier–Stokes system

In this subsection we recall the compressible Navier–Stokes system for a Newtonian fluid and introduce the several dimensionless parameters used in the next sections. The CNS reads

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0, & \rho \geq 0, \\ \frac{\partial \rho u}{\partial t} + \operatorname{div}(\rho u \otimes u) - \operatorname{div}(2\mu D(u)) - \nabla(\lambda \operatorname{div} u) + \nabla p = f, \\ \frac{\partial \rho e}{\partial t} + \operatorname{div}(\rho u e) + p \operatorname{div} u - \operatorname{div}(k \nabla T) = 2\mu |D(u)|^2 + \lambda (\operatorname{div} u)^2. \end{cases} \quad (1)$$

In the above system, t is time, div and ∇ only act in the x variable and $x \in \mathbb{R}^N$. Moreover, ρ , u , p , e and T are respectively the density, the velocity, the pressure, the internal energy by unit mass and the temperature of the fluid. Besides, μ and λ are the so-called

Lamé viscosity coefficients and satisfy the relation $\mu \geq 0$, $N\lambda + 2\mu \geq 0$. The coefficient k is the thermal conduction coefficient and satisfies $k \geq 0$. In general μ , λ and k can depend on the thermodynamical functions and their gradients. Finally, f is the force term. For geophysical flows we will consider a force which is the sum of the gravitational force and the Coriolis force, namely $f = \rho \mathbf{g} + \rho \Omega \mathbf{e} \times u$, where Ω is the rotation frequency and $\mathbf{e}$ is the direction of rotation. We also denote $g = |\mathbf{g}|$.

The system (1) can be closed by the thermodynamic state equations, namely $p = P(\rho, T)$ and $e = e(\rho, T)$. For an ideal gas, these functions are given by

$$\begin{cases} e = C_v T, \\ p = \rho R T, \end{cases} \quad (2)$$

where $R > 0$ is the ideal gas constant and $C_v > 0$ is a constant. We also define $C_p = R + C_v$. The constant C_v and C_p are respectively the specific heats at constant volume and constant pressure. We also define the adiabatic constant $\gamma = C_p / C_v$.

The system formed by (1) and (2) is closed. There is an other important thermodynamical function, namely the entropy. It is defined by the following thermodynamic relation

$$T \, dS = \frac{\partial e}{\partial T} dT + \left(\frac{\partial e}{\partial \rho} - \frac{p}{\rho^2} \right) d\rho. \quad (3)$$

For an ideal gas, (3) yields $\frac{\partial S}{\partial T} = \frac{C_v}{T}$ and $\frac{\partial S}{\partial \rho} = -\frac{R}{\rho}$. Hence S is given by $S = C_v \log(T/\rho^{\gamma-1})$. In particular, we can replace the third equation of (1) by an equation for the entropy, namely

$$\frac{\partial \rho S}{\partial t} + \operatorname{div}(\rho u S) = \frac{1}{T} \operatorname{div}(k \nabla T) + \frac{2\mu |D(u)|^2 + \lambda (\operatorname{div} u)^2}{T}. \quad (4)$$

Let us notice that if we take $\mu = \lambda = 0$ and $k = 0$ then (4) reduces to a transport equation and that if the entropy is constant initially $S = S_0$ then it remains constant at later times. In this case, $T = e^{S/c_v} \rho^{\gamma-1}$ and $p = R e^{S_0/c_v} \rho^\gamma$. This yields the compressible isentropic Euler system. An other model we will deal with is the isentropic compressible Navier–Stokes system (69). It corresponds to the case $k = 0$, S is constant and we neglect the variation of S due to the viscous effects. However, (69) cannot be rigorously derived from (1) in any asymptotic regime.

1.4. Dimensionless parameters

Let us now define the different dimensionless parameters. We take $\underline{t}$, $\underline{L}$, $\underline{U}$, $\underline{\rho}$ and $\underline{P}$ to be respectively the characteristic time scale, the characteristic length scale, the characteristic velocity scale, the characteristic density scale and the characteristic pressure scale. This means that each time or length is made dimensionless by dividing it by $\underline{t}$ or $\underline{L}$. Hence, we can define a dimensionless time and dimensionless length by $\tilde{t} = t/\underline{t}$ and $\tilde{x} = x/\underline{L}$. We can

do the same for all the other quantities. We also take characteristic values of μ and k which we denote $\underline{\mu}$ and $\underline{k}$. These are equal to μ and k if they are constant.

The Strouhal number and Reynolds number are defined by

$$St = \frac{\underline{L}}{\underline{t} \underline{U}} \quad (5)$$

and

$$Re = \frac{\underline{L} \underline{U}}{\underline{\mu} / \underline{\rho}}. \quad (6)$$

A small Strouhal number St corresponds to the longtime behavior of a system. A large Reynolds number Re corresponds to small viscous effects.

The acoustic waves propagates at the sound speed which is given in the isentropic case by $c^2 = \frac{\partial p}{\partial \rho} = \gamma RT$. Hence we can define the Mach number as the ratio between $\underline{U}$ and c , namely

$$Ma = \frac{\underline{U}}{c} = \frac{\underline{U}}{\sqrt{\gamma R \underline{t}}}. \quad (7)$$

When $Ma < 1$, we have a subsonic flow and when $Ma > 1$, we have a supersonic flow.

The velocity and the temperature satisfy both a diffusion equation with a diffusivity given respectively by $\underline{\mu} / \underline{\rho}$ and $\underline{k} / (C_v \underline{\rho})$. The ratio between this two numbers is the Prandtl number

$$Pr = \gamma \frac{C_v \underline{\mu}}{\underline{k}} = \frac{C_p \underline{\mu}}{\underline{k}}. \quad (8)$$

Now, we will introduce some other dimensionless parameters related to the gravity force and the Coriolis force. First, let us introduce a vertical length scale $\underline{H}$. Hence the gravity wave speed is given by $\sqrt{g \underline{H}}$ and we can define the Froude number which measures the importance of the gravity force. It is the ratio between $\underline{U}$ and $\sqrt{g \underline{H}}$, namely

$$Fr = \frac{\underline{U}}{\sqrt{g \underline{H}}}. \quad (9)$$

The Rossby number measures the importance of the Earth's rotation. It is the ratio between the rotation time scale $t_\Omega = 1/\Omega$ and the fluid time scale $t_U = \underline{L}/\underline{U}$. It is given by

$$Ro = \frac{\underline{U}}{\Omega \underline{L}}. \quad (10)$$

Since, we have two length scale, we can define the ratio between $\underline{H}$ and $\underline{L}$, $\delta = \underline{H}/\underline{L}$. It measures how shallow the fluid is.

In Section 2 we study the inviscid limit, namely the limit when the Reynolds number goes to infinity. We will mostly emphasis the problem of boundary layers. In Section 3 we study the compressible–incompressible limit, namely the limit when the Mach number goes to infinity and the density becomes almost constant. We also study the limit when γ (the adiabatic constant) goes to ∞ . We will emphasis the problem of oscillations in time. In Section 4 we study rotating fluid at high frequency. In Section 5 we will study the hydrodynamic limit of the Boltzmann equation and derive several compressible and incompressible fluid systems. In Section 6 we will recall few results about the homogenization of the Stokes, the Euler and the compressible Navier–Stokes system. In Section 7.1 we will give some other examples of singular limits which were not studied in the previous sections. Finally, in Section 7.2 we will give some concluding remarks.

Let us end this introduction by giving some general references about fluid mechanics. We refer to [33,119,122] for mathematical results about the incompressible Euler equation. We refer to [39,108,163] for mathematical results about the incompressible Navier–Stokes system. We refer to [66,109,117,139] for results about the compressible Navier–Stokes system. We also refer to [176,177] for many formal asymptotic developments and to [80,118,144] for physical and mathematical results about the geophysical equations.

2. The inviscid limit

The Navier–Stokes system is the basic mathematical model for viscous incompressible flows. It reads

$$\begin{cases} \partial_t u^v + u^v \cdot \nabla u^v - \nu \Delta u^v + \nabla p = 0, \\ \operatorname{div}(u^v) = 0, \\ u^v = 0 \quad \text{on } \partial\Omega, \end{cases} \quad (11)$$

where u^v is the velocity, p is the pressure and ν is the kinematic viscosity. We can define a typical length scale L and a typical velocity U . The dimensionless parameter $Re = UL/\nu$ is very important to compare the properties of different flows. When Re is very large (ν very small), we can expect that the Navier–Stokes system (NS_ν) behaves like the Euler system

$$\begin{cases} \partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p = 0, \\ \operatorname{div} \mathbf{u} = 0, \\ \mathbf{u} \cdot \mathbf{n} = 0 \quad \text{on } \partial\Omega. \end{cases} \quad (12)$$

The zero-viscosity limit for the incompressible Navier–Stokes equation in a bounded domain, with Dirichlet boundary conditions, is one of the most challenging open problems in Fluid Mechanics. This is due to the formation of a boundary layer which appears because we cannot impose a Dirichlet boundary condition for the Euler equation. This boundary layer satisfies formally the Prandtl equations, which seem to be ill-posed in general. Let us first state some results in the whole space where the boundary layer problem does not occur.

2.1. The whole space case

The inviscid limit in the whole space case was performed by several authors, we can refer for instance to Swann [159] and Kato [95]. They basically prove the following result. Take the Navier–Stokes system in the whole space $\mathbb{R}^d$,

$$\partial_t u^n + \operatorname{div}(u^n \otimes u^n) - \nu_n \Delta u^n = -\nabla p \quad \text{in } \mathbb{R}^d, \quad (13)$$

$$\operatorname{div}(u^n) = 0 \quad \text{in } \mathbb{R}^d, \quad (14)$$

$$u^n(t=0) = u_0 \quad \text{with } \operatorname{div}(u_0) = 0, \quad (15)$$

where ν_n goes to 0 when n goes to infinity.

THEOREM 2.1. *Let $s > d/2 + 1$, and $u_0 \in H^s(\mathbb{R}^d)$. If T^* is the time of existence and $u \in C_{\text{loc}}([0, T^*]; H^s)$ is the solution of the Euler system*

$$\partial_t u + \operatorname{div}(u \otimes u) = -\nabla p \quad \text{in } \mathbb{R}^d, \quad (16)$$

$$\operatorname{div} u = 0 \quad \text{in } \mathbb{R}^d, \quad (17)$$

$$u(t=0) = u_0 \quad \text{with } \operatorname{div}(u_0) = 0, \quad (18)$$

then for all $0 < T < T^$, there exists ν_0 such that for all $\nu_n \leq \nu_0$, the Navier–Stokes system (13)–(15) has a unique solution $u^n \in C([0, T]; H^s(\mathbb{R}^d))$ and for each $t \in [0, T]$, $u(t) = \lim_{n \rightarrow \infty} u^n(t)$ exists strongly in $H^s(\mathbb{R}^d)$ uniformly in $t \in [0, T]$. Moreover,*

$$\|u^n - u\|_{L^\infty(0, T; H^{s-2})} \leq C \nu_n, \quad (19)$$

where C depends only on u .

We point out that this result can be easily extended to the periodic case and more generally to domains without boundaries.

IDEA OF THE PROOF. The proof of this theorem is based on a standard Grönwall inequality (see [38, 95, 159]). Let us start by proving (19). First, we see that we can solve the Navier–Stokes system and Euler system in $C([0, T]; H^s(\mathbb{R}^d))$ on some time interval independent of ν_n with bounds which are independent of n . This is because there is no boundary. Then we can write an energy estimate in H^{s-2} for $w^n = u^n - u$,

$$\begin{aligned} & \partial_t \|w^n\|_{H^{s-2}}^2 + \nu_n \|\nabla w^n\|_{H^{s-2}}^2 \\ & \leq (C(\|u\|_{H^s} + \|w^n\|_{H^s}) \|w^n\|_{H^{s-2}} + \nu_n \|\Delta u\|_{H^{s-2}}) \|w^n\|_{H^{s-2}} \end{aligned} \quad (20)$$

and by the Grönwall lemma, we can deduce that (19) holds. It is easy to see that the above argument holds as long as we can solve the Euler system and that we can take any T such that $T < T^*$ (see [38]). Notice that in [38], the regularity required is $s - 2 > d/2 + 1$.

However, it seems that this is not necessary modulo the regularization argument given below.

Interpolating between (19) and the uniform bound for w^n in $C([0, T]; H^s(\mathbb{R}^d))$, we deduce that u^n converges to u in $H^{s'}$ for any $s' < s$ and for $s - 2 < s' < s$, we have

$$\|u^n - u\|_{L^\infty(0, T; H^{s'})} \leq C v_n^{(s-s')/2}. \quad (21)$$

To get the convergence in H^s requires a regularization of the initial data. For all $\delta > 0$, we take u_0^δ such that $\|u_0^\delta\|_{H^s} \leq C \|u_0\|_{H^s}$, $\|u_0^\delta\|_{H^{s+1}} \leq C/\delta$, $\|u_0^\delta\|_{H^{s+2}} \leq C/\delta^2$ and for some s' such that $d/2 < s' < s - 1$, we have $\|u_0^\delta - u_0\|_{H^{s'}} \leq C\delta^{s-s'}$. Such a u_0^δ can be easily constructed by taking $u_0^\delta = \mathcal{F}^{-1}(\mathbb{1}_{\{|\xi| \leq 1/\delta\}} \mathcal{F}u_0)$. Let v^δ be the solution of the Euler system (16)–(18) with the initial data $v^\delta(t=0) = u_0^\delta$. Then, setting $w^\delta = v^\delta - u$, we have

$$\partial_t \|w^\delta\|_{H^s}^2 \leq C(\|u\|_{H^s} + \|v^\delta\|_{H^s}) \|w^\delta\|_{H^s}^2 + C\|v^\delta\|_{H^{s+1}} \|w^\delta\|_{H^s} \|w^\delta\|_{L^\infty}. \quad (22)$$

Then, we notice that on some time interval $[0, T]$, $T < T^*$ (T depends only on $\|u_0\|_{H^s}$), we have $\|v^\delta\|_{H^{s+1}} \leq C/\delta$ and $\|v^\delta\|_{H^{s+2}} \leq C/\delta^2$. Moreover, writing (22) at the regularity s' , we can prove easily that $\|w^\delta\|_{L^\infty(0, T; H^{s'})} \leq C\delta^{s-s'}$. Hence, (22) gives

$$\partial_t \|w^\delta\|_{H^s} \leq C(\|u\|_{H^s} + \|v^\delta\|_{H^s}) \|w^\delta\|_{H^s} + C\delta^{s-s'-1}. \quad (23)$$

Hence w^δ goes to zero in $L^\infty(0, T; H^s)$, namely v^δ goes to v in $L^\infty(0, T; H^s)$. Writing an energy estimate for $w^{n,\delta} = u^n - v^\delta$, we get (here we drop the n and δ)

$$\begin{aligned} \partial_t \|w\|_{H^s}^2 + v_n \|\nabla w\|_{H^s}^2 \\ \leq C(\|w\|_{L^\infty} \|v^\delta\|_{H^{s+1}} \|w\|_{H^s} + (\|v^\delta\|_{H^s} + \|u^n\|_{H^s}) \|w\|_{H^s}^2) \\ + v_n \|v^\delta\|_{H^{s+2}} \|w\|_{H^s}. \end{aligned} \quad (24)$$

Hence, we get

$$\begin{aligned} \partial_t \|w\|_{H^s} \leq C\|u^n - u\|_{L^\infty} \|v^\delta\|_{H^{s+1}} + C\|v^\delta - u\|_{L^\infty} \|v^\delta\|_{H^{s+1}} \\ + v_n \|v^\delta\|_{H^{s+2}} + C(\|v^\delta\|_{H^s} + \|u^n\|_{H^s}) \|w\|_{H^s}. \end{aligned} \quad (25)$$

Since u^n converges to u in H^{s-1} , we deduce that

$$\|u^n - u\|_{L^\infty} \leq \|u^n - u\|_{H^{s-1}} \leq C(v_n)^{1/2}. \quad (26)$$

Taking $\delta = \delta_n$ such that $\delta = \delta_n$ and v_n/δ_n^2 go to zero when n goes to infinity, we deduce that

$$\partial_t \|w^{n,\delta}\|_{H^s} \leq C\left(\frac{v^{1/2}}{\delta} + \delta^{s-s'-1} + \frac{v}{\delta^2} + \|w^{n,\delta} v\|_{H^s}\right). \quad (27)$$

Hence, by the Grönwall lemma, we deduce that $w^{n,\delta}$ goes to zero in $L^\infty(0, T; H^s)$ and that u^n goes to u in $L^\infty(0, T; H^s)$. $\square$

2.1.1. The 2D case. We notice that the time T^* is related to the existence time for the Euler system (16). If $d = 2$ it is known [171, 175] that the Euler system (16) has a global solution and hence one can take any time $T < \infty$ in the above theorem.

Also in the 2D case, one can lower the regularity assumption. Indeed Yudovich [175] proved that if $\omega_0 = \text{curl}(u_0) \in L^\infty \cap L^p$ for some $1 < p < \infty$ then the Euler system (16) has a unique global solution. It was proved in [34] that the solution to the Navier–Stokes system converges in $L^\infty((0, T); L^2)$ to the solution of the Euler system if we only assume that $\omega_0 = \text{curl}(u_0) \in L^\infty \cap L^p$. More precisely, Chemin [34] proves that

$$\|u^n - u\|_{L^\infty(0, T; L^2)} \leq C \|\text{curl}(u_0)\|_{L^\infty \cap L^2}^{\frac{1}{2}} (v_n T)^{\frac{1}{2} \exp(-C \|\text{curl}(u_0)\|_{L^\infty \cap L^2} T)}. \quad (28)$$

Notice that here, the rate of convergence deteriorates with time. This does not happen if we also know that u is in $L^\infty(0, T; Lip)$ as was proved by Constantin and Wu [40].

For vortex patches, namely the case where $\text{curl}(u_0)$ is the characteristic function of a $C^{1+\alpha}$ domain $\alpha > 0$, it was proved in [32] (see also [22]) that the characteristic function of $\text{curl}(u)$ remains a $C^{1+\alpha}$ domain and that the velocity u is in $L^\infty_{\text{loc}}(\mathbb{R}; Lip)$. It was proved in [40, 41] that under the condition, $u \in L^\infty_{\text{loc}}(\mathbb{R}; Lip)$, the estimate (28) is actually better since there is no loss for the rate of convergence, namely

$$\|u^n - u\|_{L^\infty(0, T; L^2)} \leq C (v_n T)^{1/2}. \quad (29)$$

In [41] the authors also prove some estimate in L^p spaces for the difference between the vorticities, in particular they prove for $p \geq 2$ that $\|\text{curl}(u^n - u)\|_{L^\infty(0, T; L^p)} \leq C v_n^{1/4 p - \varepsilon}$ for some short time T and $\varepsilon > 0$.

Concerning vortex patches one can give more precise results about the convergence. It was proved by Danchin [42] that the boundary of the patch under the Navier–Stokes flow converges to the boundary of the patch under the Euler flow. A similar result is also proved in higher dimension locally in time [43]. Also, in [1] a better rate of convergence is given for vortex patches, namely

$$\|u^n - v\|_{L^\infty(0, T; L^2)} \leq C (v_n T)^{3/4} \quad (30)$$

which is optimal (see also [129] for a similar result in 3D).

Let us end this subsection by the vortex sheet case, namely the case where the vorticity is a measure. For the 2D case, it is known that we have existence of weak solutions for the Euler system if we only assume that $u_0 \in L^2$ and $\omega_0 \in L^1 \cap L^p$, $1 < p$. In this case, extracting a subsequence, we can prove the weak convergence of the solutions to the Navier–Stokes system toward a weak solution to the Euler system. Indeed, from the bound we have on the vorticity $\text{curl}(u^n) \in L^\infty(0, T; L^p)$, we deduce that u^n is bounded in

$L^\infty(0, T; W^{1,p})$ and since $\partial_t u^n$ is bounded in $L^\infty(0, T; H^{-1})$ we deduce that u^n is pre-compact in $L^2 L^2_{\text{loc}}$. Then extracting a subsequence, we deduce that u^n converges to some u and u is a weak solution of the Euler system. Here, the main point is that $W^{1,p}(\mathbb{R}^2)$ is compactly injected in $L^2_{\text{loc}}(\mathbb{R}^2)$. The above argument does not work if $p = 1$. However, the best result in this direction is due to Delort [50] where he can prove the weak convergence under the assumption that the initial vorticity is compactly supported, belongs to $H^{-1}(\mathbb{R}^2)$ and can be decomposed into two parts: one being a nonnegative measure, the other belonging to some $L^q(\mathbb{R}^2)$, $q > 1$. The proof requires a precise analysis to rule out concentrations at the limit.

2.2. The case of the Dirichlet boundary condition

Let us consider the limit from (11) toward (12). In the region close to the boundary the length scale becomes very small and we can not neglect the viscous effect. In 1904, Prandtl [145] suggested that there exists a thin layer called boundary layer, where the solution u^ν undergoes a sharp transition from a solution to the Euler system to the no-slip boundary condition $u^\nu = 0$ on $\partial\Omega$ of the Navier–Stokes system. In other words, Prandtl proves formally that $u^\nu = \mathbf{u} + u^\nu_{\text{BL}}$ where u^ν_{BL} is small except near the boundary. Giving a rigorous justification of this formal expansion is still an open problem. We refer to [150, 151] for a justification in the analytic case.

There are many review papers about the inviscid limit of the Navier–Stokes in a bounded domain and the Prandtl system (see [29,60]). We also refer to [83] for a review about boundary layers.

2.2.1. Formal derivation of Prandtl system. To illustrate this, we consider a two-dimensional (planar) flow $u^\nu = (u, v)$ in the half-space $\{(x, y) \mid y > 0\}$ subject to the following initial condition $u^\nu(t=0, x, y) = u_0^\nu(x, y)$, boundary condition $u^\nu(t, x, y=0) = 0$ and $u^\nu \rightarrow (U_0, 0)$ when $y \rightarrow \infty$. Taking the typical length and velocity of order one, the Reynolds number reduces to $Re = \nu^{-1}$. Let $\varepsilon = Re^{-1/2} = \sqrt{\nu}$. Near the boundary, the Euler system is not a good approximation. We introduce new independent variables and new unknowns

$$\begin{aligned} \tilde{t} &= t, & \tilde{x} &= x, & \tilde{y} &= \frac{y}{\varepsilon}, \\ (\tilde{u}, \tilde{v})(\tilde{t}, \tilde{x}, \tilde{y}) &= \left(u, \frac{v}{\varepsilon}\right)(\tilde{t}, \tilde{x}, \varepsilon\tilde{y}). \end{aligned}$$

Notice that when $\tilde{y}$ is of order one, $y = \varepsilon\tilde{y}$ is of order ε . Rewriting the Navier–Stokes system in terms of the new variables and unknowns yields

$$\begin{cases} \tilde{u}_{\tilde{t}} + \tilde{u}\tilde{u}_{\tilde{x}} + \tilde{v}\tilde{u}_{\tilde{y}} - \tilde{u}_{\tilde{y}\tilde{y}} - \varepsilon^2\tilde{u}_{\tilde{x}\tilde{x}} + p_{\tilde{x}} = 0, \\ \varepsilon^2(\tilde{v}_{\tilde{t}} + \tilde{u}\tilde{v}_{\tilde{x}} + \tilde{v}\tilde{v}_{\tilde{y}} - \tilde{v}_{\tilde{y}\tilde{y}}) - \varepsilon^4\tilde{v}_{\tilde{x}\tilde{x}} + p_{\tilde{y}} = 0, \\ \tilde{u}_{\tilde{x}} + \tilde{v}_{\tilde{y}} = 0. \end{cases}$$

Neglecting the terms of order ε^2 and ε^4 yields

$$\begin{cases} \tilde{u}_{\tilde{t}} + \tilde{u}\tilde{u}_{\tilde{x}} + \tilde{v}\tilde{u}_{\tilde{y}} - \tilde{u}_{\tilde{y}\tilde{y}} + p_{\tilde{x}} = 0, \\ p_{\tilde{y}} = 0, \quad \tilde{u}_{\tilde{x}} + \tilde{v}_{\tilde{y}} = 0. \end{cases}$$

Since p does not depend on $\tilde{y}$, we deduce that the pressure does not vary within the boundary layer and can be recovered from the Euler system (12) when $y = 0$, namely $p_x(t, x) = -(U_t + UU_x)(t, x, y = 0)$, since $V(t, x, y = 0) = 0$. Going back to the old variables, we obtain

$$\begin{cases} u_t + uu_x + vu_y - vu_{yy} + p_x = 0, \\ u_x + v_y = 0, \end{cases} \quad (31)$$

which is the so-called Prandtl system. It should be supplemented with the following boundary conditions

$$\begin{cases} u(t, x, y = 0) = v(t, x, y = 0) = 0, \\ u(t, x, y) \rightarrow U(t, x, 0) \quad \text{as } y \rightarrow \infty. \end{cases} \quad (32)$$

Formally, a good approximation of u^ν should be $\mathbf{u} + u_{\text{BL}}^\nu$, where $\mathbf{u}$ is the solution of the Euler system (12) and $\mathbf{u}(t, x, 0) + u_{\text{BL}}^\nu$ is the solution of the Prandtl system (31), (32).

Replacing the Navier–Stokes system by the Euler system in the interior and the Prandtl system near the boundary requires a justification. Mathematically this can be formulated as a convergence theorem when ν goes to 0, namely $u^\nu - (\mathbf{u} + u_{\text{BL}}^\nu)$ goes to 0 when ν goes to 0 in L^∞ or in some energy space. In its whole generality this is still a major open problem in fluid mechanics. This is due to problems related to the well-posedness of the Prandtl system. Indeed, under some monotonicity condition on the initial data, Oleinik proved the local existence for the Prandtl system [140,141] (see also [142]). These solutions can be extended as global weak solutions [173]. However, E and Engquist [61] proved a blow up result for the Prandtl system for some special type of initial data. For general initial data, it is not known whether we have local well-posedness or not. Moreover, even if we have existence for Prandtl system there are other problems related to the instability of some solutions to the Prandtl system [82] which may prevent the convergence.

2.2.2. The analytic case. In this subsection we will present the result of [150,151]. We will just give an informal statement since the result requires the definition of several spaces to keep track of the analyticity of the solution.

THEOREM 2.2. *Suppose that $\mathbf{u}(t, x, y)$ and $\mathbf{u}(t, x, 0) + u_{\text{BL}}^\nu$ are respectively the solutions of the Euler system (12) and the Prandtl system (31), (32) which are analytic in the space variables. Then for a short time independent of $\sqrt{\nu}$, there is an analytic solution u of the Navier–Stokes equations such that it is given by $u = \mathbf{u} + O(\sqrt{\nu})$ in the interior and $u = \mathbf{u}(t, x, 0) + u_{\text{BL}}^\nu + O(\sqrt{\nu})$ inside the boundary layer.*

We refer to [29] for a sketch of the proof and to [150,151] for the complete proof.

2.2.3. Kato's criterion of convergence. The convergence of $u^\nu - (\mathbf{u} + u_{\text{BL}}^\nu)$ to 0 when ν goes to 0 in L^2 is still an open problem. Kato [96] gave a very simple criterion which is equivalent to the convergence of u^ν to $\mathbf{u}$ in L^2 .

First let us notice that working with strong solutions to the Navier–Stokes system does not really help. Indeed, the existence of strong solution for $d \geq 3$ only holds on a time interval $[0, T_\nu]$ where T_ν may go to zero when ν goes to 0. Also, for $d = 2$, working with strong solutions does not help since the higher Sobolev norms blow up when ν goes to zero. This is why we consider a family of weak solutions u^ν to the Navier–Stokes system (11) with an initial data u_0^ν . We assume that $u^\nu \in C_w([0, T]; L^2(\Omega)) \cap L^2(0, T; H^1(\Omega))$ for all $T > 0$, $\operatorname{div} u^\nu = 0$ and (11) holds in the sense of distributions, namely

$$-\int_{\Omega} u_0^\nu \phi(t=0) + \int_0^T \int_{\Omega} -u^\nu \otimes u^\nu : \nabla \phi + \nu \nabla u^\nu \cdot \nabla \phi - u^\nu \partial_t \phi \, dx \, dt = 0 \quad (33)$$

for all $\phi \in C_0^\infty([0, T] \times \Omega)$, $\operatorname{div} \phi = 0$ and the following energy inequality holds

$$\int_{\Omega} |u^\nu(t, x)|^2 \, dx + 2\nu \int_0^T \int_{\Omega} |\nabla u^\nu(s, x)|^2 \, dx \, ds \leq \int_{\Omega} |u_0^\nu(x)|^2 \, dx. \quad (34)$$

Assume that u_0^ν is divergence-free and converges in L^2 to some u_0 and $u_0 \in H^s$, $s > d/2 + 1$. Let $\mathbf{u}$ be the unique strong solution of the Euler system (12) with the initial data u_0 in the space $C([0, T^*]; H^s)$ for some $T^* \leq \infty$ and $T^* = \infty$ if $d = 2$. We refer to [162] and [33] for this existence result. Kato proves the following convergence criterion.

THEOREM 2.3. *For $0 < T < T^*$, the following conditions are equivalent:*

- (i) $u^\nu(t, \cdot)$ converges to $\mathbf{u}(t, \cdot)$ in $L^2(\Omega)$ uniformly for $t \in [0, T]$;
- (ii)

$$\nu \int_0^T \int_{\Gamma_\nu} |\nabla u^\nu|^2 \, dx \, dt \rightarrow 0 \quad (35)$$

when ν goes to 0. Here Γ_ν is a strip of width $O(\nu)$ around the boundary $\partial\Omega$;

- (iii)

$$\nu \int_0^T \int_{\Omega} |\nabla u^\nu|^2 \, dx \, dt \rightarrow 0. \quad (36)$$

IDEA OF THE PROOF. We will just give a sketch of the proof of (ii) implies (i). The idea is to construct a corrector or boundary layer which allows to recover the Dirichlet boundary condition for the difference $u^\nu - \mathbf{u}$ and which satisfies some natural bounds. Kato constructs such a corrector $\mathcal{B}^\nu$ which is divergence free and with a support contained in a strip of size $O(\nu)$ around $\partial\Omega$. Then, considering $u^\nu - \mathbf{u} - \mathcal{B}$, he can write the following energy

estimate where $\mathbf{u}^{\mathcal{B}} = \mathbf{u} + \mathcal{B}^v$,

$$\begin{aligned} & \frac{1}{2} \|u^v - \mathbf{u}\|_{L^2}^2(t) + \nu \int_0^t \|\nabla u^v\|_{L^2}^2 \, ds \\ & \leq \int_0^t \int_{\Omega} -(u^v \times u^v) : \nabla \mathbf{u}^{\mathcal{B}} + \mathbf{u} \cdot \nabla \mathbf{u} \cdot u^v + \nu \nabla u^v \cdot \nabla \mathbf{u}^{\mathcal{B}} \, dx \, ds + o(1) \end{aligned} \quad (37)$$

for $0 < t \leq T$, where $o(1)$ goes to zero when ν goes to 0. This yields

$$\begin{aligned} & \frac{1}{2} \|u^v - \mathbf{u}\|_{L^2}^2(t) + \nu \int_0^t \|\nabla u^v\|_{L^2}^2 \\ & \leq \int_0^t \int_{\Omega} -(u^v - \mathbf{u}) \times (u^v - \mathbf{u}) : \nabla \mathbf{u} - u^v \times u^v : \nabla \mathcal{B}^v \\ & \quad + \int_0^t \int_{\Omega} \nu \nabla u^v \cdot \nabla \mathbf{u}^{\mathcal{B}} \, dx \, dt + o(1). \end{aligned} \quad (38)$$

Then, using some natural L^2 and L^∞ bounds satisfied by $\mathcal{B}^v$, the Hardy–Littlewood inequality for the second term on the right-hand side of (38) and applying a Gronwall lemma, Kato gets

$$\|u^v - \mathbf{u}\|_{L^2}^2(t) \leq \int_0^t K \|u^v - \mathbf{u}\|_{L^2}^2 + R(s) \, ds + o(1) \quad (39)$$

for some constant K related to the L^∞ norm of $\nabla \mathbf{u}$ and

$$R(t) \leq K \int_0^t \nu \|\nabla u^v\|_{L^2(\Gamma_v)}^2 + K \nu \|\nabla u^v\|_{L^2} + K \nu^{1/2} \|\nabla u^v\|_{L^2(\Gamma_v)}.$$

This ends the proof of the uniform convergence in L^2 . Notice that it also proves (iii) since the total dissipation appears on the left-hand side of (38). $\square$

In the same spirit as the Kato criterion, Temam and Wang [164] give a different criterion based on the magnitude of the pressure at the boundary. They prove that if there exists some $0 \leq \delta < 1/2$ such that

$$\text{either } \nu^\delta \int_0^T \|p^v\|_{H^{1/2}(\partial\Omega)} \leq C \quad \text{or} \quad \nu^{\delta+1/4} \int_0^T \|\nabla p^v\|_{L^2(\partial\Omega)} \leq C \quad (40)$$

then the convergence of u^v toward $\mathbf{u}$ holds and

$$\|u^v - \mathbf{u}\|_{L^2} \leq C \nu^{(1-2\delta)/5}. \quad (41)$$

Also, in [169] Wang gives a criterion which only involves the tangential derivative of the velocity, namely $\nabla_\tau u^v$. However, he needs a control on a strip of size bigger than ν .

Concerning bounded domain with boundary conditions other than the Dirichlet boundary condition, let us mention that in [165], Temam and Wang prove the convergence of the solutions to the Navier–Stokes system toward a solution of the Euler system in the non-characteristic case, namely the normal velocity is prescribed at the boundary. In this case a boundary layer of size ν can be constructed.

Let us also mention that in [11] Bardos treats the case of a bounded domain with a boundary condition on the vorticity, which does not engender any boundary layer. He has a result similar to Theorem 2.1.

Also, in [36], the vanishing viscosity limit is considered with the Navier (friction) boundary condition.

2.2.4. Different vertical and horizontal viscosities. One of the main ideas of Kato in the previous subsection is to take the freedom of using a corrector which does not necessary satisfy the Prandtl system. The same idea was used in [123] to get a complete convergence result without any condition on the dissipation in the case we take different vertical and horizontal viscosities. We consider the following system of equations $(NS_{\nu,\eta})$

$$\partial_t u^n + \operatorname{div}(u^n \otimes u^n) - \nu \partial_z^2 u^n - \eta \Delta_{x,y} u^n = -\nabla p \quad \text{in } \Omega, \quad (42)$$

$$\operatorname{div}(u^n) = 0 \quad \text{in } \Omega, \quad (43)$$

$$u^n = 0 \quad \text{in } \partial\Omega, \quad (44)$$

$$u^n(0) = u_0^n \quad \text{with } \nabla \cdot u_0^n = 0, \quad (45)$$

where $\Omega = \omega \times (0, h)$ or $\Omega = \omega \times (0, \infty)$ and $\omega = \mathbb{T}^2$ or $\mathbb{R}^2$, $\nu = \nu_n$, $\eta = \eta_n$. We want to point out here that this anisotropy is classical in geophysical flows. In fact instead of putting the classical viscosity $-\bar{\nu} \Delta$ of the fluid in the equation, meteorologists often model turbulent diffusion by putting a viscosity of the form $-A_H \Delta_{x,y} - A_V \partial_z^2$, where A_H and A_V are empiric constants, and where A_V is usually much smaller than A_H . (For instance in the ocean, A_V ranges from 1 to $10^3 \text{ cm}^2/\text{s}$ whereas A_H ranges from 10^5 to $10^8 \text{ cm}^2/\text{s}$.) We recall that the viscosity of the water is of order $10^{-2} \text{ cm}^2/\text{s}$.) We refer to the book of Pedlovsky [144], Chapter 4, for a more complete discussion. When η, ν go to 0, we expect that u^n converges to the solution of the Euler system

$$\begin{cases} \partial_t w + \operatorname{div}(w \otimes w) = -\nabla p & \text{in } \Omega, \\ \operatorname{div} w = 0 & \text{in } \Omega, \\ w \cdot n = \pm w_3 = 0 & \text{on } \partial\Omega, \\ w(t=0) = w^0. \end{cases} \quad (46)$$

It turns out that we are able to justify this formal derivation under an additional condition on the ratio of the vertical and horizontal viscosities.

THEOREM 2.4. *Let $s > 5/2$, and*

$$w^0 \in H^s(\Omega)^3, \quad \operatorname{div}(w^0) = 0, \quad w^0 \cdot n = 0 \quad \text{on } \partial\Omega.$$

We assume that $u^n(0)$ converges in $L^2(\Omega)$, to w^0 and $v, \eta, v/\eta$ go to 0, then any sequence of global weak solutions (à la Leray) u^n of (42)–(45) satisfying the energy inequality satisfies

$$\begin{aligned} u^n - w &\rightarrow 0 \quad \text{in } L_{\text{loc}}^\infty([0, T^*]; L^2(\Omega)), \\ \sqrt{\eta} \nabla_{x,y} u^n, \sqrt{v} \partial_z u^n &\rightarrow 0 \quad \text{in } L_{\text{loc}}^2([0, T^*]; L^2(\Omega)), \end{aligned}$$

where w is the unique solution of (46) in $C([0, T^*]; H^s(\Omega)^3)$.

We give here a sketch of the proof and refer to [123] for a complete proof. The existence of global weak solutions for $(NS_{v,\eta})$, satisfying the energy inequality is due to Leray [102–104] (see also [90] and [39,163] for some references about weak solutions of the Navier–Stokes)

$$\frac{1}{2} \|u^n(t)\|_{L^2}^2 + \nu \int_0^t \|\partial_z u^n\|_{L^2}^2 ds + \eta \int_0^t \|\partial_x u^n\|_{L^2}^2 + \|\partial_y u^n\|_{L^2}^2 \leq \frac{1}{2} \|u_0^n\|_{L^2}^2. \quad (47)$$

This estimate does not show that u^n is bounded in $L^2(0, T; H^1)$ and hence if we extract a subsequence still denoted by u^n converging weakly to u in $L^\infty(0, T; L^2)$, we cannot deduce that $u^n \otimes u^n$ converges weakly to $w \otimes w$. If we try to use energy estimates to show that $u^n - w$ remains small we see that the integrations by parts introduce terms that we cannot control, since $u^n - w$ does not vanish at the boundary. Hence, we must construct a boundary layer which allows us to recover the Dirichlet boundary conditions. Hence, $\mathcal{B}^n$ will be a corrector of small L^2 norm, and localized near $\partial\Omega$ (we take here the case where $\Omega = \omega \times (0, \infty)$ not to deal with boundary conditions near $z = h$)

$$\begin{cases} \mathcal{B}^n(z=0) + w(z=0) = 0, & \mathcal{B}^n(z=\infty) = 0, \\ \operatorname{div}(\mathcal{B}^n) = 0, & \mathcal{B}^n \rightarrow 0 \text{ in } L_{\text{loc}}^\infty([0, T^*]; L^2) \end{cases}$$

a possible choice is to take $\mathcal{B}^n$ of the form

$$\mathcal{B}^n = -w(z=0)e^{-z/\sqrt{\nu\zeta}} + \dots,$$

where ζ is a free parameter to be chosen later. We want to explain now the idea of the proof. Instead of using energy estimates on $u^n - w$, we will work with $v^n = u^n - (w + \mathcal{B}^n)$. Next we write the following equation satisfied by $w^{\mathcal{B}} = w + \mathcal{B}^n$ (in what follows, we will write $\mathcal{B}$ instead of $\mathcal{B}^n$)

$$\begin{aligned} \partial_t w^{\mathcal{B}} + w^{\mathcal{B}} \cdot \nabla w^{\mathcal{B}} - \nu \partial_z^2 w^{\mathcal{B}} - \eta \Delta_{x,y} w^{\mathcal{B}} \\ = \partial_t \mathcal{B} + \mathcal{B} \cdot \nabla w^{\mathcal{B}} + w \cdot \nabla \mathcal{B} - \nu \partial_z^2 w^{\mathcal{B}} - \eta \Delta_{x,y} w^{\mathcal{B}} - \nabla p \end{aligned} \quad (48)$$

which yields the following energy equality

$$\begin{aligned} & \frac{1}{2} \|w^{\mathcal{B}}(t)\|_{L^2}^2 + \nu \int_0^t \|\partial_z w^{\mathcal{B}}(s)\|_{L^2}^2 ds + \eta \int_0^t \|\partial_x w^{\mathcal{B}}\|_{L^2}^2 + \|\partial_y w^{\mathcal{B}}\|_{L^2}^2 \\ &= \frac{1}{2} \|w^{\mathcal{B}}(0)\|_{L^2}^2 + \int_0^t w^{\mathcal{B}} \cdot [\partial_t \mathcal{B} + w \cdot \nabla \mathcal{B} - \nu \partial_z^2 w^{\mathcal{B}} - \eta \Delta_{x,y} w^{\mathcal{B}}]. \end{aligned} \quad (49)$$

Next, using the weak formulation of (42), we get for all t ,

$$\begin{aligned} & \int_{\Omega} u^n \cdot w^{\mathcal{B}}(t) + \nu \int_0^t \int_{\Omega} \partial_z w^{\mathcal{B}}(s) u^n + \eta \int_0^t \int_{\Omega} \partial_x w^{\mathcal{B}} \partial_x u^n + \partial_y w^{\mathcal{B}} \partial_y u^n \\ &= \int_{\Omega} u^n \cdot w^{\mathcal{B}}(0) + \int_0^t \int_{\Omega} u^n \cdot \nabla w^{\mathcal{B}} u^n \\ &+ u^n \cdot [\partial_t \mathcal{B} - w \cdot \nabla w - \nu \partial_z^2 w^{\mathcal{B}} - \eta \Delta_{x,y} w^{\mathcal{B}}]. \end{aligned} \quad (50)$$

Then adding up (47) and (49) and subtracting (50), we get

$$\begin{aligned} & \frac{1}{2} \|v(t)\|_{L^2}^2 + \nu \int_0^t \|\partial_z v\|_{L^2}^2 ds + \eta \int_0^t \|\partial_x v\|_{L^2}^2 + \|\partial_y v\|_{L^2}^2 \\ & \frac{1}{2} \|v_0\|_{L^2}^2 + \int_0^t \int_{\Omega} v \cdot [\partial_t \mathcal{B} - \nu \partial_z^2 w^{\mathcal{B}} - \eta \Delta_{x,y} w^{\mathcal{B}}] \\ &+ w \cdot \nabla \mathcal{B} w^{\mathcal{B}} - u^n \cdot \nabla w^{\mathcal{B}} u^n + w \cdot \nabla w u^n. \end{aligned} \quad (51)$$

Finally, using that $\int (u \cdot \nabla q) q = 0$, we get

$$\begin{aligned} & \int_{\Omega} w \cdot \nabla \mathcal{B} w^{\mathcal{B}} - u^n \cdot \nabla w^{\mathcal{B}} u^n + w \cdot \nabla w u^n \\ &= \int_{\Omega} -w^{\mathcal{B}} \cdot \nabla \mathcal{B} v - \mathcal{B} \cdot \nabla w v - v \cdot \nabla w^{\mathcal{B}} v. \end{aligned}$$

Now, we want to use a Gronwall lemma to deduce that $\|v(t)\|_{L^2}^2$ remains small. By studying two terms among those occurring in the right-hand side of the energy estimate (51), we want to show why we need the condition $\nu/\eta \rightarrow 0$. In fact,

$$\begin{aligned} \left| \int_{\Omega} v_3 \partial_z \mathcal{B} v \right| &\leq \int \frac{v_3}{z} z^2 \partial_z \mathcal{B} \frac{v}{z} \\ &\leq C \|\partial_z v_3\|_{L^2} \sqrt{\nu \zeta} \|w\|_{L^\infty} \|\partial_z v\|_{L^2} \\ &\leq C \zeta \|\partial_z v_3\|_{L^2}^2 \|w\|_{L^\infty}^2 + \frac{\nu}{4} \|\partial_z v\|_{L^2}^2, \end{aligned}$$

where we have used the divergence-free condition $\partial_z v_3 = -\partial_x v_1 - \partial_y v_2$. We see from this

term that we need the following condition to absorb the first term by the viscosity in (51): $C\zeta\|w\|_{L^\infty}^2 \leq \eta$. On the other hand, the second term can be treated as follows

$$\begin{aligned} \left| v \int_{\Omega} \partial_z^2 \mathcal{B} v \right| &\leq v \|\partial_z v\|_{L^2} \|\partial_z \mathcal{B}\|_{L^2} \\ &\leq \frac{v}{4} \|\partial_z v\|_{L^2}^2 + v \|\partial_z \mathcal{B}\|_{L^2}^2 \\ &\leq \frac{v}{4} \|\partial_z v\|_{L^2}^2 + v \|w\|_{L^\infty}^2 \frac{1}{\sqrt{v\zeta}}. \end{aligned}$$

The second term on the right-hand side must go to zero, this is the case if we have $v/\zeta \rightarrow 0$. Finally, we see that

$$\text{if } \frac{v}{\eta} \rightarrow 0 \quad \text{then } \zeta = \frac{\eta}{C\|w\|_{L^\infty}^2}$$

is a possible choice.

2.3. Weak limit

We want to conclude this section by mentioning an other important question in the inviscid limit of the Navier–Stokes even in the case without boundary. Consider any sequence of weak solutions to the Navier–Stokes system with viscosity ν . What can we say about this sequence when ν goes to 0. In Section 2.1 we saw that if the initial data is regular enough then the sequence converges to the solution of the Euler system on some small time interval. Moreover, in the 2D case, we can take initial data such that the vorticity is a signed measure and still prove that the solutions of the Navier–Stokes system weakly converge to a solution of the Euler system [50]. Can we say more? What can we say if we only assume that $u_0 \in L^2$? We mention here two attempts to explain what happens based on two notions of “very weak” solutions to the Euler system.

2.3.1. Measure valued solutions. In their three papers [57–59] Diperna and Majda studied the behavior of sequences of approximate solutions to the Euler system. In the introduction of [59], they state “a sequence of Leray–Hopf weak solutions of the Navier–Stokes equations converges in the high Reynolds number limit to a measure-valued solution of Euler defined for all positive times”. They introduced the following notion of measure valued solutions to the Euler system.

DEFINITION 2.5. Let $\mathcal{O}$ be a smooth domain of $\mathbb{R}^d$, μ a nonnegative measure of $\mathcal{M}(\mathcal{O})$ and $(t, x) \rightarrow (v_{(t,x)}^1, v_{(t,x)}^2)$ a $dt d\mu$ -measurable map from $(0, T) \times \mathcal{O}$ to $\mathcal{M}^+(\mathbb{R}^d) \times \text{Prob}(\mathbb{S}^{d-1})$. We also denote $\mu = \mu_s + f dt dx$ the Lebesgue decomposition of μ into its singular and absolutely continuous parts. Then the triple (μ, v^1, v^2) is called a measure

valued solution of the incompressible Euler system if

$$\begin{aligned} \operatorname{div} \left[\left\langle v_{(t,x)}^1, \frac{v}{1+|v|^2} \right\rangle (1+f) \right] &= 0 \quad \text{and} \\ \int \int \phi_t \cdot \left\langle v_{(t,x)}^1, \frac{v}{1+|v|^2} \right\rangle (1+f) \, dt \, dx + \nabla \phi : \left\langle v_{(t,x)}^2, \xi \times \xi \right\rangle d\mu &= 0 \end{aligned} \quad (52)$$

for all smooth divergence-free vector field $\phi(t, x)$.

Of course a weak solution u of the Euler system defines a measure valued solution by taking $f = \mu = |u|^2$, $v_{(t,x)}^1 = \delta_{v=u(t,x)}$ and $v_{(t,x)}^2(\xi) = \delta_{\xi=u/|u|}$ if $u(t, x) \neq 0$.

They also define the notion of generalized Young measure for a sequence $\{v^\varepsilon\}$ bounded in $L^2(\mathcal{O})$.

THEOREM 2.6. *If $\{v^\varepsilon\}$ is an arbitrary family of functions whose L^2 norm on a set $\mathcal{O}$ is uniformly bounded, then extracting a subsequence, there exist a measure $\mu \in \mathcal{M}(\mathcal{O})$ such that*

$$|v_\varepsilon|^2 \rightarrow \mu \quad \text{in } \mathcal{M}(\mathcal{O}), \quad (53)$$

and a μ -measurable map $x \rightarrow (v_{(x)}^1, v_{(x)}^2)$ from $\mathcal{O}$ to $\mathcal{M}^+(\mathbb{R}^d) \times \operatorname{Prob}(\mathbb{S}^{d-1})$ such that for all

$$g(v) = g_0(v)(1+|v|^2) + g_H\left(\frac{v}{|v|}\right)|v|^2,$$

where g_0 lies in the space $C_0(\mathbb{R}^d)$ of continuous function vanishing at infinity and g_H lies in the space $C(\mathbb{S}^{d-1})$ of continuous function on the unit sphere, we have

$$g(v_\varepsilon) \rightarrow \langle v_{(x)}^1, g_0(v) \rangle (1+f) \, dx + \langle v_{(x)}^2, g_H(v) \rangle d\mu \quad \text{in } \mathcal{D}', \quad (54)$$

where f denotes the Radon–Nikodym derivative of μ with respect to dx . The triple (μ, v^1, v^2) is called the generalized Young measure of the sequence $\{v^\varepsilon\}$.

The notion of generalized Young measure can be extended to the case the function v_ε also depend on t . The above two definitions are linked by the following theorem.

THEOREM 2.7. *Assume v_ε is a sequence of functions satisfying $\operatorname{div}(v_\varepsilon) = 0$, v_ε is bounded in $L^2((0, T) \times \mathcal{O})$ and for all divergence-free test function ϕ in $C_0^\infty((0, T) \times \mathcal{O})$,*

$$\lim_{\varepsilon \rightarrow 0} \int \int (\phi_t \cdot v_\varepsilon + \nabla \phi : v_\varepsilon \times v_\varepsilon) \, dt \, dx = 0. \quad (55)$$

Then, if (μ, v^1, v^2) is a generalized Young measure of the sequence $\{v^\varepsilon\}$ then it defines a measure-valued solution to the Euler system.

Of course, one of the main application of this theorem is the case where v_ε satisfies the Navier–Stokes equation with a vanishing viscosity since it implies (55).

2.3.2. Dissipative solutions. An other notion of “very weak” solutions to the Euler system was introduced by Lions [108]. As stated by Lions, it is not clear whether this notion is relevant. Its only merits are the fact that such solutions exist and are global and as long as a “smooth” solution exists with the same initial data, any such dissipative solution coincides with it. Let us point out that such a uniqueness property does not hold for the measure-valued solutions of the previous subsection. Before defining dissipative solutions, let us introduce few notations. For a divergence-free smooth test function v of $[0, \infty) \times \mathbb{R}^d$, we define

$$E(v) = -\frac{\partial v}{\partial t} - P(v \cdot \nabla v), \quad (56)$$

where P is the Leray projector on divergence free vector fields. We also denote $d(v)_{ij} = \frac{1}{2}(\partial_i v_j + \partial_j v_i)$, the symmetric part of ∇v . For $t \geq 0$, let

$$\|d^-\|_\infty = \left\| \sup_{|\xi|=1} -(d\xi, \xi)_+ \right\|_{L^\infty(\mathbb{R}^d)}. \quad (57)$$

DEFINITION 2.8. Let $u \in L^\infty(0, \infty; L^2) \cap C([0, \infty); L^2_w)$. Then u is a dissipative solution of the Euler system

$$\begin{cases} \partial_t u + \operatorname{div}(u \otimes u) = -\nabla p & \text{in } \mathbb{R}^d, \\ \operatorname{div} u = 0 & \text{in } \mathbb{R}^d, \\ u(t=0) = u^0, \end{cases} \quad (58)$$

if $u(0) = u^0$, $\operatorname{div} u = 0$ and for all divergence-free smooth test function v , we have

$$\begin{aligned} \|(u - v)(t)\|_{L^2(\mathbb{R}^d)}^2 &\leq e^{2\int_0^t \|d^-\|_\infty} \|(u - v)(0)\|_{L^2(\mathbb{R}^d)}^2 \\ &\quad + 2 \int_0^t \int_{\mathbb{R}^d} e^{2\int_s^t \|d^-\|_\infty} E(v) \cdot (u - v) \, ds. \end{aligned} \quad (59)$$

In [108] Lions proves the following result.

THEOREM 2.9. Let u^ν be a sequence of Leray-weak solutions to the Navier–Stokes system with viscosity ν and initial data u_0^ν . In particular, it satisfies

$$\frac{d}{dt} \|u^\nu\|_{L^2(\mathbb{R}^d)}^2 + \nu \|\nabla u^\nu\|_{L^2(\mathbb{R}^d)}^2 \leq 0 \quad \text{in } \mathcal{D}', \quad (60)$$

$u^\nu \in L^2(0, T; H^1) \cap L^\infty(0, \infty; L^2) \cap C([0, \infty); L^2_w)$ for all $T > 0$ and $u^\nu(t)$ goes to u_0^ν in $L^2(\mathbb{R}^d)$ when t goes to 0. Assume that u_0^ν converges in L^2 to u^0 then, extracting a

subsequence, u^ν converges weakly-* in $L^\infty(0, \infty; L^2)$ to some u and converges weakly in L^2 uniformly in $t \in [0, T]$ to u . Moreover, u is a dissipative solution of the Euler system.

Let us give a sketch of the proof. From (60), we can deduce that for all divergence-free test function v , we have

$$\begin{aligned} & \frac{d}{dt} \|u^\nu - v\|_{L^2(\mathbb{R}^d)}^2 + \nu \|\nabla u^\nu\|_{L^2(\mathbb{R}^d)}^2 \\ & \leq 2 \|d^-\|_\infty \|u^\nu - v\|_{L^2(\mathbb{R}^d)}^2 \\ & \quad + 2 \int E(v) \cdot (u^\nu - v) dx + C(v) \nu \|\nabla u^\nu\|_{L^2}. \end{aligned} \quad (61)$$

Then, we can apply a Grönwall lemma to get

$$\begin{aligned} & \| (u^\nu - v)(t) \|_{L^2(\mathbb{R}^d)}^2 \\ & \leq e^{2 \int_0^t \|d^-\|_\infty} \| (u_0^\nu - v(0)) \|_{L^2(\mathbb{R}^d)}^2 \\ & \quad + 2 \int_0^t \int_{\mathbb{R}^d} e^{2 \int_s^t \|d^-\|_\infty} E(v) \cdot (u^\nu - v) ds + C_T(v) \nu. \end{aligned} \quad (62)$$

Then, we can extract a subsequence of u^ν which converges weakly-* in $L^\infty(0, \infty; L^2)$. Passing to the limit in (62), we deduce that u is a dissipative solution of the Euler system.

3. Compressible–incompressible limit

It is well known from a Fluid Mechanics viewpoint that one can derive formally incompressible models such as the Incompressible Navier–Stokes system or the Euler system from compressible ones namely compressible Navier–Stokes system (CNS) when the Mach number goes to 0 and the density becomes constant. There are several mathematical justifications of this derivation. One can put these works in two categories depending on the type of solutions considered. Indeed, one viewpoint consists on looking at local strong solutions and trying to prove existence on some time interval independent of the Mach number and then studying the limit when the Mach number goes to zero. This was initiated by Klainerman and Majda [97] (see also Ebin [62]). The second point of view consists on retrieving the Leray global weak solutions [103,104] of the incompressible Navier–Stokes system starting from global weak solutions of the compressible Navier–Stokes system (see [111]). Let us also mention that there were many works about this limit during the last 10 years and that there are many review papers about it (see for instance [47,70,124,155]).

3.1. Formal limit

We first wish to recall the general set up for such asymptotic problems. We will present it for the compressible isentropic Navier–Stokes system. The unknowns $(\tilde{\rho}, v)$ are respectively the density and the velocity of the fluid (gas) and solve on $(0, \infty) \times \mathbb{R}^N$,

$$\frac{\partial \tilde{\rho}}{\partial t} + \operatorname{div}(\tilde{\rho}v) = 0, \quad \tilde{\rho} \geq 0 \quad (63)$$

$$\frac{\partial \tilde{\rho}v}{\partial t} + \operatorname{div}(\tilde{\rho}v \otimes v) - \tilde{\mu} \Delta v - \tilde{\xi} \nabla \operatorname{div} v + \nabla \tilde{p} = 0 \quad (64)$$

and

$$\tilde{p} = a \tilde{\rho}^\gamma, \quad (65)$$

where $N \geq 2$, $\tilde{\mu} > 0$, $\tilde{\mu} + \tilde{\xi} > 0$, $a > 0$ and $\gamma > 1$ are given.

From a physical view-point, the fluid should behave (asymptotically) like an incompressible one when the density is almost constant, the velocity is small and we look at large time scales. More precisely, we scale ρ and v (and thus p) in the following way

$$\tilde{\rho} = \rho(\varepsilon t, x), \quad v = \varepsilon u(\varepsilon t, x) \quad (66)$$

and we assume that the viscosity coefficients μ, ξ are also small and scale like

$$\tilde{\mu} = \varepsilon \mu_\varepsilon, \quad \tilde{\xi} = \varepsilon \xi_\varepsilon, \quad (67)$$

where $\varepsilon \in (0, 1)$ is a “small parameter” and the normalized coefficient $\mu_\varepsilon, \xi_\varepsilon$ satisfy

$$\mu_\varepsilon \rightarrow \mu, \quad \mu_\varepsilon \rightarrow \xi \quad \text{as } \varepsilon \text{ goes to } 0_+. \quad (68)$$

We shall always assume that we have either $\mu > 0$ and $\mu + \xi > 0$ or $\mu = 0$.

With the preceding scalings, the system (63)–(65) yields

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0, & \rho \geq 0, \\ \frac{\partial \rho u}{\partial t} + \operatorname{div}(\rho u \otimes u) - \mu_\varepsilon \Delta u - \xi_\varepsilon \nabla \operatorname{div} u + \frac{a}{\varepsilon^2} \nabla \rho^\gamma = 0. \end{cases} \quad (69)$$

We may now explain the heuristics which lead to incompressible models. First of all, the second equation (for the momentum ρu) indicates that ρ should be like $\bar{\rho} + O(\varepsilon^2)$ where $\bar{\rho}$ is a constant. Of course, $\bar{\rho} \geq 0$ and we always assume that $\bar{\rho} > 0$ (in order to avoid the trivial case $\bar{\rho} = 0$). Obviously, we need to assume this property holds initially (at $t = 0$). And, let us also remark that by a simple (multiplicative) scaling, we may always assume without loss of generality that $\bar{\rho} = 1$.

Since ρ goes to 1, we expect that the first equation in (69) yields at the limit: $\operatorname{div} u = 0$. And writing $\nabla \rho^\gamma = \nabla(\rho^\gamma - 1)$, we deduce from the second equation in (69) that we have in the case when $\mu > 0$

$$\frac{\partial u}{\partial t} + \operatorname{div}(u \otimes u) - \mu \Delta u + \nabla \pi = 0 \quad (70)$$

or when $\mu = 0$

$$\frac{\partial u}{\partial t} + \operatorname{div}(u \otimes u) + \nabla \pi = 0, \quad (71)$$

where π is the “limit” of $(\rho^\gamma - 1)/\varepsilon^2$. In other words, we recover the incompressible Navier–Stokes equations (70) or the incompressible Euler equations (71), and the hydrostatic pressure appears as the limit of the “renormalized” thermodynamical pressure $(\rho^\gamma - 1)/\varepsilon^2$. In fact, as we shall see later on, the derivation of (70) (or (71)) is basically correct even globally in time, for global weak solutions; but the limiting process for the pressure is much more involved and may, depending on the initial conditions, incorporate additional terms coming from the oscillations in $\operatorname{div}(\rho_\varepsilon u_\varepsilon \otimes u_\varepsilon)$.

This section about the compressible incompressible limit is organized as follows. In the next Section 3.2 we recall the results of Klainerman and Majda [97,98] for the strong solutions to the isentropic compressible Navier–Stokes when the Mach number goes to zero. Then, we give several extensions of that result by taking general or “ill-prepared” initial data [154,166]. Also we state result about long time existence for the slightly compressible system [87,89]. We also present results in “almost” critical spaces [45,46]. In Section 3.3 we recall the results of convergence from the global weak solutions to the isentropic compressible Navier–Stokes toward the global weak solutions of the incompressible Navier–Stokes. In the last Section 3.5 we state some newer results about the nonisentropic case [133,134].

We will not mention result about the steady problem and refer to [17,106,111].

3.2. The case of strong solutions

The first mathematical justification of the incompressible limit is due to Ebin [62]. By using Lagrangian coordinates and a geometric description of the equations, he proved that “slightly compressible fluid motion can be described as a motion with a strong constraining force, while incompressible fluid flow is the analogous constrained motion”. The first justification using PDE methods was done by Klainerman and Majda [97,98] using the theory of singular limits of symmetric hyperbolic systems. We should also mention the work of Kreiss [99] about problems with different time scales but which requires the control of more time derivatives at time $t = 0$.

We consider the compressible Euler system which can be recovered from (69), by taking $\mu_\varepsilon = \xi_\varepsilon = 0$,

$$\begin{cases} \frac{\partial \rho_\varepsilon}{\partial t} + u_\varepsilon \cdot \nabla \rho_\varepsilon + \rho_\varepsilon \operatorname{div} u_\varepsilon = 0, & \rho_\varepsilon \geq 0, \\ \rho_\varepsilon \left(\frac{\partial u_\varepsilon}{\partial t} + u_\varepsilon \cdot \nabla u_\varepsilon \right) + \frac{1}{\varepsilon^2} \nabla p_\varepsilon = 0, \end{cases} \quad (72)$$

where p_ε and ρ_ε are related by $p_\varepsilon = a \rho_\varepsilon^\gamma$ where $a > 0$ and $\gamma \geq 1$ are given constants. They consider the above system in the torus or the whole space, $\Omega = \mathbb{T}^N$ or $\Omega = \mathbb{R}^N$ with the following initial data

$$u_\varepsilon(t=0, x) = u_\varepsilon^0(x), \quad p_\varepsilon(t=0, x) = p_\varepsilon^0(x). \quad (73)$$

Notice that we can retrieve the initial data for ρ_ε from the initial data for p_ε . Here $\|\cdot\|_s$ will denote the H^s norm and $s_0 = [\frac{N}{2}] + 1$.

THEOREM 3.1. *Assume the initial data (73) satisfies*

$$\|u_\varepsilon^0(x)\|_s + \frac{1}{\varepsilon} \|p_\varepsilon^0(x) - \underline{p}\|_s \leq C_0 \quad (74)$$

for some constants $\underline{p} > 0$ and C_0 and some $s \geq s_0 + 1$. Then there exists an ε_0 and a fixed time interval $[0, T]$ with T depending only upon $\|u_\varepsilon^0(x)\|_{s_0+1} + \frac{1}{\varepsilon} \|p_\varepsilon^0(x) - p_0\|_{s_0+1}$ and a constant C_s such that for $\varepsilon < \varepsilon_0$, a classical solution of the compressible Euler system exists on $[0, T] \times \Omega$ and satisfies

$$\sup_{0 \leq t \leq T} \|u_\varepsilon\|_s + \frac{1}{\varepsilon} \|p_\varepsilon^0 - \underline{p}\|_s + \varepsilon \left\| \frac{\partial u_\varepsilon}{\partial t} \right\|_{s-1} + \left\| \frac{\partial p_\varepsilon}{\partial t} \right\|_{s-1} \leq C_s. \quad (75)$$

Moreover, if the initial data satisfies the additional condition

$$\begin{aligned} u_\varepsilon^0(x) &= u^0(x) + \varepsilon u^1(x), \quad \operatorname{div} u^0 = 0, \\ p_\varepsilon^0(x) &= \underline{p} + \varepsilon^2 p^1(x), \\ \|u^1(x)\|_s + \|p^1(x)\|_s &\leq C_0, \end{aligned} \quad (76)$$

then, on the same time interval $[0, T]$, we have

$$\sup_{0 \leq t \leq T} \left\| \frac{\partial u_\varepsilon}{\partial t} \right\|_{s-1} + \varepsilon^{-1} \left\| \frac{\partial p_\varepsilon}{\partial t} \right\|_{s-1} \leq C_s^1 \quad (77)$$

and as ε goes to 0, u_ε converges weakly in $L^\infty([0, T]; H^s)$ and uniformly in $C_{\text{loc}}([0, T] \times \Omega)$ to u^∞ where u^∞ satisfies the incompressible Euler system

$$\begin{cases} \frac{\partial u^\infty}{\partial t} + u^\infty \cdot \nabla u^\infty + \nabla p^\infty = 0 \\ u^\infty(t=0, x) = u^0(x), \quad \operatorname{div} u^\infty = 0. \end{cases} \quad (78)$$

The condition (76) means that the flow is initially almost incompressible and that the density is initially almost constant. These data are called “well-prepared” initial data. The more general condition (74) will be called general initial data or “ill-prepared” initial data. Notice that we still need to assume that $p_\varepsilon^0 - \underline{p}$ is of order ε this is because, we need to make a change a variable $q_\varepsilon = \varepsilon^{-1}(p_\varepsilon - \underline{p})$ to write our system in a form which is suitable for energy estimates, we will denote $q_\varepsilon^0 = \varepsilon^{-1}(p_\varepsilon^0 - \underline{p})$.

IDEA OF THE PROOF. We rewrite the system in terms of the new unknowns $(u_\varepsilon, q_\varepsilon)$ where $q_\varepsilon = \varepsilon^{-1}(p_\varepsilon - \underline{p})$

$$\begin{cases} \frac{\partial q_\varepsilon}{\partial t} + u_\varepsilon \cdot \nabla q_\varepsilon + \frac{\gamma}{\varepsilon}(\underline{p} + \varepsilon q_\varepsilon) \operatorname{div} u_\varepsilon = 0, & \rho_\varepsilon \geq 0, \\ \frac{\partial u_\varepsilon}{\partial t} + u_\varepsilon \cdot \nabla u_\varepsilon + \frac{1}{\varepsilon(\underline{p} + \varepsilon q_\varepsilon)^{1/\gamma}} \nabla q_\varepsilon = 0. \end{cases} \quad (79)$$

To prove (75), we just need to prove H^s estimates on some time interval $[0, T]$ which is independent of ε . For each ε , we denote

$$E_s(t) = \int \sum_{|\alpha|=s} \frac{1}{(\underline{p} + \varepsilon q_\varepsilon)^{1/\gamma}} |\partial^\alpha q_\varepsilon|^2 + \gamma(\underline{p} + \varepsilon q_\varepsilon) |\partial^\alpha u_\varepsilon|^2. \quad (80)$$

Then, we can prove that $\partial_t E_s \leq C(E_s)^2$ where C does not depend on $\varepsilon < \varepsilon_0$. This shows that there exists a time of existence T which is uniform in ε .

Next, we have to prove (77) and the convergence toward the incompressible system (78) under the well-prepared condition (76). We notice that, taking the time derivative of (79), we can write a hyperbolic equation for $(\partial_t u_\varepsilon, \partial_t q_\varepsilon)$ which is similar to (79). To prove uniform bounds for $(\partial_t u_\varepsilon, \partial_t q_\varepsilon)$ in H^{s-1} on some time interval $[0, T]$ we only need to have bounds in H^{s-1} initially. This follows immediately from (76). Hence, if (76) holds then (77) holds. Moreover, by simple compactness arguments, we can extract a subsequence such that $(u_\varepsilon, q_\varepsilon)$ converges in $C([0, T]; H_{\text{loc}}^{s-\kappa})$ to some (u, q) for $\kappa > 0$. Then, it is easy to see that u satisfies the Euler system (78) by passing weakly to the limit in the different terms. Since, we have uniqueness for (78), we deduce the convergence of the whole sequence. $\square$

REMARK 3.2. 1. In [97,98], the authors also deal with the Navier–Stokes case by proving that the viscosity does not affect the leading hyperbolic behavior.

2. For the “well-prepared” case, the convergence stated in the theorem can be improved to a convergence in $C([0, T]; H^s)$ (see [18,19]).

During the last 25 years there were different extensions of this result in different directions. First, there were results trying to take more general initial data. These results require some analysis of the acoustic waves. Then, there were results about more general models, namely the nonisentropic model (the entropy is not constant and is transported by the flow). Also, there were results trying to improve the minimum regularity required for the convergence.

3.2.1. General initial data. In the whole space $\mathbb{R}^N$ (see [166]) or in the exterior of a bounded domain (see [92,93]), the result of [98] has been extended to the case of general initial data or “ill-prepared” initial data. The convergence toward the incompressible limit holds locally in space. However, we do not have uniform convergence near $t = 0$ due to the presence of an initial layer in time. This layer comes from acoustic waves that go to infinity. We have the following result.

THEOREM 3.3 ($\Omega = \mathbb{R}^N$). *Assume the initial data (73) satisfies (74) and that $(u_\varepsilon^0(x), q_\varepsilon^0(x))$ converges to some $(u^0(x), q^0(x))$ in H^s , then the solution constructed in Theorem 3.1 satisfies*

$$(q_\varepsilon, u_\varepsilon) \rightarrow (0, u^\infty) \quad (81)$$

weakly in $L^\infty((0, T); H^s)$ and strongly in $C_{\text{loc}}^0((0, T] \times \mathbb{R}^N)$ where u^∞ is the unique solution to the incompressible Euler system (78) with the initial data Pu^0 where P is the Leray projection onto divergence free vector fields $P = \text{Id} - \nabla \Delta^{-1} \nabla \cdot$.*

In the periodic case $\mathbb{T}^N$, Schochet [154] extends the result of [98] to the case of “ill-prepared” initial data. He proves the same Theorem 3.3 in the periodic case with the only difference that the $(0, u^\infty)$ is replaced by (c, u^∞) for some constant c and that the convergence is only weak due to the acoustic waves. The convergence is strong for the divergence-free part Pu_ε .

THEOREM 3.4 ($\Omega = \mathbb{T}^N$). *Assume the initial data (73) satisfies (74) and that $(u_\varepsilon^0(x), q_\varepsilon^0(x))$ converges to some $(u^0(x), q^0(x))$ in H^s , then the solution constructed in Theorem 3.1 satisfies*

$$(q_\varepsilon, u_\varepsilon) \rightarrow (c, u^\infty) \quad (82)$$

weakly in $L^\infty((0, T); H^s)$ where u^∞ is the unique solution to the incompressible Euler system (78) with the initial data Pu^0 where P is the Leray projection onto divergence free vector fields $P = \text{Id} - \nabla \Delta^{-1} \nabla \cdot$. Moreover, Pu_ε converges strongly in $C_{\text{loc}}^0([0, T] \times \mathbb{T}^N)$ to u^∞ .*

IDEA OF THE PROOFS. The idea of Theorem 3.4 is to use the group method to filter the oscillations. We also would like to mention that ideas close to the group method were also developed by Joly, Métivier and Rauch [94]. We introduce the following group $(\mathcal{L}(\tau), \tau \in \mathbb{R})$ defined by $e^{\tau L}$ where L is the operator defined on $\mathcal{D}' \times (\mathcal{D}')^N$, by

$$L \begin{pmatrix} \varphi \\ v \end{pmatrix} = - \begin{pmatrix} \gamma \underline{p} \operatorname{div} v \\ \frac{1}{\underline{p}^{1/\gamma}} \nabla \varphi \end{pmatrix}. \quad (83)$$

It is easy to check that $e^{\tau L}$ is an isometry on each $H^s \times (H^s)^N$ for all $s \in \mathbb{R}$ and for all τ . This show that if we define $\begin{pmatrix} \varphi(\tau) \\ v(\tau) \end{pmatrix} = e^{\tau L} \begin{pmatrix} \varphi_0 \\ v_0 \end{pmatrix}$ then it solves

$$\frac{\partial \varphi}{\partial \tau} = -\gamma \underline{p} \operatorname{div} v, \quad \frac{\partial v}{\partial \tau} = -\frac{1}{\underline{p}^{1/\gamma}} \nabla \varphi.$$

If we denote $U_\varepsilon = \mathfrak{U}(q_\varepsilon, u_\varepsilon)$, then $V_\varepsilon = \mathcal{L}(-t/\varepsilon)U_\varepsilon$ is such that $\partial_t V_\varepsilon$ is bounded in $L^\infty(0, T; H^{s-1})$. Then, we can use compactness argument to extract a subsequence which converges to some V in $C([0, T]; H_{\text{loc}}^{s-\kappa})$. Now, passing to the limit in the equation satisfied by V requires the study of resonances. It turns out these resonances do not affect the divergence-free flow. See also Section 3.3.5 for more about resonances.

If we consider the whole space case, we notice that the long time behavior of the operator $e^{\tau L}$ is not the same in the whole space and in the torus. Indeed, in the whole space we have dispersion and the following Strichartz [158] type estimate holds

$$\left\| e^{\frac{t}{\varepsilon} L} \begin{pmatrix} \psi \\ \nabla \phi \end{pmatrix} \right\|_{L^p(\mathbb{R}; W^{s,q}(\mathbb{R}^N))} \leq C \varepsilon^{1/p} \left\| \begin{pmatrix} \psi \\ \nabla \phi \end{pmatrix} \right\|_{H^{s+\sigma}} \quad (84)$$

for all $p, q > 2$ and $\sigma > 0$ such that

$$\frac{2}{q} = (N-1) \left(\frac{1}{2} - \frac{1}{p} \right), \quad \sigma = \frac{1}{2} + \frac{1}{p} - \frac{1}{q}.$$

This dispersion allows for the convergence in $C_{\text{loc}}^0((0, T) \times \mathbb{R}^N)$ (see also [52]). $\square$

3.2.2. Long time existence for the compressible system. In [87] Hagstrom and Lorenz give a result about the global existence of strong solutions to the slightly compressible Navier–Stokes system in 2D for initial data which are close to the incompressible, namely satisfying a condition of the type (76). Also, in [89] Hoff gives a similar result in dimension 2 or 3 with a force term under some assumptions about the limit system. These two results use different properties of the system. However, they both use in a critical way the presence of the viscosity. Consider the system (69) with $a\gamma = 1$, $\mu_\varepsilon = \mu > 0$, $\xi_\varepsilon = \xi$ and $\mu + \xi > 0$. The limit system reads

$$\begin{cases} \frac{\partial u^\infty}{\partial t} + u^\infty \cdot \nabla u^\infty - \mu u^\infty + \nabla p^\infty = 0, \\ u^\infty(t=0, x) = u^0(x), \quad \operatorname{div} u^\infty = 0. \end{cases} \quad (85)$$

In [87] the following result is proved.

THEOREM 3.5. *Let $u^0 \in C^\infty(\mathbb{T}^2)$ be an incompressible velocity field and $\pi_0(x) = p^\infty(t=0, x)$ where (u^∞, p^∞) is the solution to (85), $-\Delta \pi_0(x) = \sum_{i,j=1}^2 \partial_i u_j^0 \partial_j u_i^0$. There exists $\varepsilon_0 = \varepsilon_0(u^0, \mu, \xi)$ and $\delta_0 = \delta_0(u^0, \mu, \xi)$ such that if $0 < \varepsilon < \varepsilon_0$ and the initial data $(\rho_\varepsilon^0, u_\varepsilon^0)$ for (69) satisfies*

$$\|u_\varepsilon^0(x) - u^0\|_3 + \varepsilon^{-1} \|\rho_\varepsilon^0(x) - 1 - \varepsilon^2 \pi_0\|_3 \leq \delta_0, \quad (86)$$

then there exists a global solution $(\rho_\varepsilon, u_\varepsilon) \in C^\infty([0, \infty) \times \mathbb{T}^2)$ to (69) which locally converges to $(1, u^\infty)$ when ε goes to zero.

We also refer to Gallagher [69] for a similar result.

IDEA OF THE PROOF. We write $u^\varepsilon = u^\infty + u'$ and $\rho^\varepsilon = 1 + \varepsilon^2(\pi^\infty + \rho')$. Then we denote

$$w = \begin{pmatrix} u' \\ \varepsilon \rho' \end{pmatrix}.$$

Hence, w satisfies the following equation

$$w_t + (u^\infty + u') \cdot \nabla w = A_\varepsilon w + G, \quad (87)$$

where A_ε is a constant coefficient operator given by

$$A_\varepsilon = -\frac{1}{\varepsilon} \begin{pmatrix} 0 & 0 & \partial_x \\ 0 & 0 & \partial_y \\ \partial_x & \partial_y & 0 \end{pmatrix} + \begin{pmatrix} \mu \Delta + \xi \partial_{xx}^2 & \xi \partial_{xy}^2 & 0 \\ \xi \partial_{xy}^2 & \mu \Delta + \xi \partial_{yy}^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (88)$$

and G consists of nonlinear terms involving (u^∞, π^∞) and w . It turns out that this term can be controlled for long time due to the exponential decay of the incompressible Navier–Stokes solution u^∞ .

Equation (87) is a coupled parabolic–hyperbolic system where the large hyperbolic part is symmetric. Even though (87) is not completely parabolic, in particular there is no viscosity in the third equation, the coupling between the three equations yields some decay for w . This cannot be seen from the standard L^2 estimate but requires the use of a different scalar product. We denote $\widehat{A}_\varepsilon(k)$ the symbol of A_ε , $k \in \mathbb{Z}^2$ which can be obtained from A_ε by replacing ∂_x by ik_1 and ∂_y by ik_2 . Then a symmetrizer $H(k)$ can be constructed for (87) satisfying the following lemma.

LEMMA 3.6 [87]. *There exist c_0, c_1, C_1, C_2 depending on μ, ξ, ε_0 such that for $0 < \varepsilon < \varepsilon_0$ there are Hermitian matrices $H(k, \varepsilon) \in \mathbb{C}^{3 \times 3}$ satisfying*

$$\begin{aligned} 0 &< (I - C_1 \varepsilon I) \leq H \leq (I + C_1 \varepsilon I), \\ q^* (H \widehat{A}_\varepsilon(k) + \widehat{A}_\varepsilon(k)^* H) q \\ &\leq -c_0 q^* H q - c_1 |k|^2 (|q_1|^2 + |q_2|^2) \quad \forall q \in \mathbb{C}^3, \\ |H - I| &\leq \frac{C_2 \varepsilon}{|k|}. \end{aligned} \quad (89)$$

Using this lemma, we can define a new inner product on $L^2(\mathbb{T}^2, \mathbb{R}^3)$ by

$$(w_1, w_2)_H = \sum_{k \in \mathbb{Z}^2} \widehat{w}_1(k)^* H(k, \varepsilon) \widehat{w}_2(k)$$

which is used to prove the exponential decay. $\square$

In [89] Hoff takes an other approach to prove the long time existence for the slightly compressible Navier–Stokes. He uses the effective viscous flux F given by

$$F = (\mu + \xi) \operatorname{div} u_\varepsilon - \varepsilon^{-2} [\rho_\varepsilon^\gamma - 1] \quad (90)$$

which satisfies the following elliptic equation

$$\Delta F = \operatorname{div}(\rho_\varepsilon \partial_t u_\varepsilon + \rho_\varepsilon u_\varepsilon \cdot \nabla u_\varepsilon - \rho_\varepsilon f), \quad (91)$$

where f is the force term. It turns out that this equation yields some regularity for F which is not shared by $\operatorname{div} u_\varepsilon$ or by $\varepsilon^{-2} [\rho_\varepsilon^\gamma - 1]$. Then, Hoff uses the equation for the density to deduce that

$$(\mu + \xi) \partial_t (\rho_\varepsilon - 1) + \varepsilon^{-2} [\rho_\varepsilon^\gamma - 1] = -\rho_\varepsilon F \quad (92)$$

from which we can deduce some decay for $(\rho_\varepsilon - 1)$ if we have some good control on F . We refer to [89] for more details.

3.2.3. Convergence in critical spaces. The compressible Navier–Stokes system (69) is invariant, up to a change of the pressure law, under the transformation

$$(\rho(t, x), u(t, x)) \rightarrow (\rho(l^2 t, lx), lu(l^2 t, lx)), \quad (93)$$

$$P(\rho) \rightarrow l^2 P(\rho). \quad (94)$$

Hence it seems natural to consider initial data $(\rho^0, u^0) \in \dot{H}^{d/2} \times H^{d/2-1}$. For fixed ε the local existence for (69) in the critical Besov space $B_{2,1}^{d/2} \times B_{2,1}^{d/2-1}$ was performed by Danchin [44]. He also proves global existence if the data is small. We refer to [44] for the precise definition of the Besov space $B_{2,1}^{d/2}$. We only recall that unlike $H^{d/2}$, $B_{2,1}^{d/2}$ is injected in L^∞ .

In [46] and [45] Danchin proves the convergence of the solutions constructed in [44] toward solutions of the incompressible Navier–Stokes system. More precisely for the critical case, namely $B_{2,1}^{d/2} \times B_{2,1}^{d/2-1}$ he proves a global existence and convergence result but only for small data. For large data he works with spaces which are slightly more regular, namely $B_{2,1}^{d/2+\kappa} \times B_{2,1}^{d/2-1+\kappa}$ or the Sobolev spaces with the same regularity. Moreover, he proves the convergence toward the incompressible Navier–Stokes system as long as the solution of the limit system exists.

3.3. The case of global weak solutions

Global weak solutions to the isentropic Navier–Stokes system were constructed by Lions [109] (see also [64] and [139]). We also refer to [65] for a review paper about the isentropic

Navier–Stokes system and to [66] for the existence of weak solutions to the full compressible system. In this subsection, we would like to study the behavior of the weak solutions constructed in [109] when the Mach number goes to zero. The first paper treating this question is [111]. In [111] the group method was used to pass to the limit in the nonlinear term. This yields the convergence in the periodic case. The result of [111] was then extended in [54] and [52] to deal with the case of a bounded domain or the whole space case. In [54] the presence of a boundary layer is responsible of the damping of the acoustic waves. In [52] the dispersion of the acoustic waves yields the local strong convergence toward the incompressible solution.

In the next Section 3.3.1, we will present in some details the simple result of [113] where the convergence is proved locally in space. This proof is independent of the boundary condition. In particular it also holds for the exterior domain.

3.3.1. The local method. Let Ω be an open bounded set in $\mathbb{R}^N$. For $\varepsilon \in (0, 1]$, we consider $(\rho_\varepsilon, u_\varepsilon)$ a weak solution of

$$\begin{cases} \frac{\partial \rho_\varepsilon}{\partial t} + \operatorname{div}(\rho_\varepsilon u_\varepsilon) = 0, & \rho_\varepsilon \geq 0, \\ \frac{\partial \rho_\varepsilon u_\varepsilon}{\partial t} + \operatorname{div}(\rho_\varepsilon u_\varepsilon \otimes u_\varepsilon) - \mu \Delta u_\varepsilon - \xi \nabla \operatorname{div} u_\varepsilon + \frac{a}{\varepsilon^2} \nabla \rho_\varepsilon^\gamma = 0 \end{cases} \quad (95)$$

in $(0, T) \times \Omega$, $T > 0$, $a > 0$, $\gamma > N/2$, $\mu > 0$ and $\mu + \xi > 0$. We assume that $\rho_\varepsilon \in L^\infty(0, T; L^\gamma) \cap C([0, T]; L^1)$, $\rho_\varepsilon |u_\varepsilon|^2 \in L^\infty(0, T; L^1)$, $u_\varepsilon \in L^2(0, T; H^1)$ and that the total energy is bounded, namely

$$\begin{cases} \int_\Omega \rho_\varepsilon |u_\varepsilon|^2 + \frac{1}{\varepsilon^2} [\rho_\varepsilon^\gamma - \bar{\rho}_\varepsilon^\gamma - \gamma \bar{\rho}_\varepsilon^{\gamma-1} (\rho_\varepsilon - \bar{\rho}_\varepsilon)] dx \leq C, & \text{a.e. } t \in (0, T), \\ \int_0^T dt \int_\Omega dx |Du_\varepsilon|^2 \leq C \end{cases} \quad (96)$$

for some positive constant C independent of ε , where $\bar{\rho}_\varepsilon$ is a positive constant such that $\bar{\rho}_\varepsilon$ and $1/\bar{\rho}_\varepsilon$ are bounded independently of ε .

We denote ρ_ε^0 and m_ε^0 the initial conditions for ρ_ε and $\rho_\varepsilon u_\varepsilon$. We also assume that $|m_\varepsilon^0|^2/\rho_\varepsilon^0$, m_ε^0 , ρ_ε^0 are bounded in L^1 , $L^{2\gamma/(\gamma+1)}$, L^γ , respectively. Extracting subsequences, we can assume that ρ_ε , $\rho_\varepsilon u_\varepsilon$, $\sqrt{\rho_\varepsilon} u_\varepsilon$, u_ε , ρ_ε^0 , m_ε^0 , $m_\varepsilon^0/\sqrt{\rho_\varepsilon^0}$ converge weakly when ε goes to zero 0, toward ρ , m , w , u , ρ^0 , m^0 , $\tilde{u}^0$ (respectively in $L^\infty(0, T; L^\gamma) - w*$, $L^\infty(0, T; L^{2\gamma/(\gamma+1)}) - w*$, $L^\infty(0, T; L^2) - w*$, $L^2(0, T; H^1)$, L^γ , $L^{2\gamma/(\gamma+1)}$, L^2) and that $\bar{\rho}_\varepsilon$ converges toward $\bar{\rho}$. Finally, we denote $V_0 = \{u \in L^2(\Omega), \int_\Omega u \varphi dx = 0 \ \forall \varphi \in C_0^\infty(\Omega), \operatorname{div} \varphi = 0 \text{ in } \Omega\}$ (if Ω is regular, then $V_0 = \{\nabla p, p \in H^1(\Omega)\}$).

The main result of [113] is the following theorem.

THEOREM 3.7. *Under the above conditions*

- (i) ρ_ε converges to $\bar{\rho}$ in $L^\infty(0, T; L^\gamma)$, and $m \equiv \sqrt{\bar{\rho}} w \equiv \bar{\rho} u$.
- (ii) The weak limit u is a solution of the incompressible Navier–Stokes system

$$\begin{cases} \frac{\partial u}{\partial t} + \operatorname{div}(u \otimes u) - \nu \Delta u + \nabla \pi = 0, & \operatorname{div} u = 0 \text{ in } \Omega \times (0, T), \\ u(t = 0, x) = u^0(x) \end{cases} \quad (97)$$

with $u \in L^2(0, T; H^1) \cap L^\infty(0, T; L^2)$, $\pi \in \mathcal{D}'$ and $v = \mu/\bar{\rho}$ and $u^0 \in \tilde{u}^0 + V_0$.

REMARK 3.8. 1. For the existence of solutions to the compressible Navier–Stokes satisfying the conditions stated above, we refer to [109].

2. Theorem 3.7 does not say anything about the boundary condition satisfied by u . This is natural since there is no boundary condition for the initial system (95). This is the reason we have an initial condition $u^0 \in \tilde{u}^0 + V_0$ which may seem vague. However, if we fix some boundary conditions, then u^0 will be completely determined. This will be done in the next subsections.

IDEA OF THE PROOF. To simplify the proof, we assume that $\bar{\rho}_\varepsilon$ goes to $\bar{\rho} = 1$.

Convergence of ρ_ε to 1. We claim that ρ_ε converges to 1 in $C([0, \infty); L^\gamma)$: indeed, for ε small enough, $\bar{\rho}_\varepsilon \in (\frac{1}{2}, \frac{3}{2})$ and thus, for all $\delta > 0$, there exists some $v_\delta > 0$ such that

$$x^\gamma + (\gamma - 1)(\bar{\rho}_\varepsilon)^\gamma - \gamma x(\bar{\rho}_\varepsilon)^{\gamma-1} \geq v_\delta |x - \bar{\rho}_\varepsilon|^\gamma \quad \text{if } |x - \bar{\rho}_\varepsilon| \geq \delta, \quad x \geq 0.$$

Hence,

$$\begin{aligned} \sup_{t \geq 0} \int |\rho_\varepsilon - 1|^\gamma &\leq \delta^\gamma |\Omega| + \sup_{t \geq 0} \left[\int \mathbb{1}_{(|\rho_\varepsilon - 1| \geq \delta)} |\rho_\varepsilon - \bar{\rho}_\varepsilon|^\gamma \right] + C |\bar{\rho}_\varepsilon - 1|^\gamma \\ &\leq \delta^\gamma |\Omega| + \frac{C \varepsilon^2}{v_\delta} + C |\bar{\rho}_\varepsilon - 1|^\gamma \end{aligned}$$

and we conclude upon letting first ε go to 0 and then δ go to 0. Actually, we need more information about this convergence and more precisely, denoting $\varphi_\varepsilon = (\rho_\varepsilon - \bar{\rho}_\varepsilon)/\varepsilon$ we can prove using some convexity inequalities that φ_ε is bounded in $L^\infty(0, T; L^2)$ if $\gamma \geq 2$. If $\gamma < 2$, then $\varphi_\varepsilon \mathbb{1}_{(|\rho_\varepsilon - 1| \leq 1/2)}$ is bounded in $L^\infty(0, T; L^2)$ and $\|\varphi_\varepsilon \mathbb{1}_{(|\rho_\varepsilon - 1| > 1/2)}\|_{L^\infty(0, T; L^\gamma)} \leq C \varepsilon^{2/\gamma-1}$.

Next, we notice that $\tau_\varepsilon = \rho_\varepsilon u_\varepsilon \otimes u_\varepsilon$ is bounded in $L^\infty(0, T; L^1) \cap L^2(0, T; L^q)$ with $1/q = 1/\gamma + (N - 2)/(2N)$ if $N \geq 3$, $1 \leq q < \gamma$ if $N = 2$. Extracting a subsequence, we denote by τ a weak limit of τ_ε . Passing to the limit in the first equation of (95), we deduce that $u \in L^2(0, T; H^1)$ satisfies $\operatorname{div} u = 0$ in $\Omega \times]0, T[$. Passing to the limit in the second equation of (95), we get

$$\frac{\partial u}{\partial t} + \operatorname{div} \tau - \mu \Delta u + \nabla \pi_1 = 0, \quad (98)$$

where $\pi_1 \in \mathcal{D}'(\Omega \times (0, T))$. We just need to prove that $\operatorname{div} \tau = \operatorname{div}(u \times u) + \nabla \pi_2$. It turns out that in general π_2 does not vanish.

Convergence of u_ε in the regular case. First, we assume that φ_ε , $\pi_\varepsilon = a\varepsilon^2(\rho_\varepsilon^\gamma - \bar{\rho}_\varepsilon^\gamma - \gamma(\rho_\varepsilon - \bar{\rho}_\varepsilon))$, $m_\varepsilon = \rho_\varepsilon u_\varepsilon$, u_ε are regular in x , uniformly in ε , i.e. φ_ε , π_ε and m_ε , are bounded

in $L^\infty(0, T; H^s)$ and u_ε is bounded in $L^2(0, T; H^s)$ for all $s \geq 0$. Next, we want to show that

$$\operatorname{div}(\rho_\varepsilon u_\varepsilon \otimes u_\varepsilon) \rightharpoonup_\varepsilon \operatorname{div}(u \otimes u) + \nabla \pi_2 \quad (99)$$

for some distribution π_2 . To this end, we will pass to the limit locally in x when ε goes to 0. Let B be a ball in our domain Ω . We want to prove the convergence stated in (99) locally in $B \times (0, T)$. We introduce the orthogonal projections P and Q defined on $L^2(B)$ by $I = P + Q$; $\operatorname{div} Pu = 0$, $\operatorname{curl}(Qu) = 0$ in B ; $Pu \cdot n = 0$ on ∂B where n stands for the exterior normal to ∂B .

Applying P to the second equation of (95), we deduce easily that $\frac{\partial}{\partial t} Pm_\varepsilon$ is bounded in $L^\infty(0, T; H^s)$ ($\forall s \geq 0$) and hence that Pm_ε converges to Pu in $C([0, T]; H^s)$ ($\forall s \geq 0$). Here, we have used that the injection of $H^r(B)$ in $H^s(B)$ is compact since B is bounded. We also deduce that Pu_ε converges to Pu in $L^2(0, T; H^s)$ since $P(u_\varepsilon - u) = P((1 - \rho_\varepsilon)u_\varepsilon) + P(\rho_\varepsilon u_\varepsilon - u)$.

Next, we decompose in B , m_ε in $u + P(m_\varepsilon - u) + Q(m_\varepsilon - u)$ and u_ε in $u + P(u_\varepsilon - u) + Q(u_\varepsilon - u)$. Hence, we can decompose in $\mathcal{D}'(B)$ $\operatorname{div}(\rho_\varepsilon u_\varepsilon \otimes u_\varepsilon)$ in 8 different terms and it is easy to see that it is sufficient to show that $\operatorname{div}(Q(m_\varepsilon - u) \otimes Q(u_\varepsilon - u))$ converges to some gradient. Moreover, since $Q(m_\varepsilon - u)$ and $Q(u_\varepsilon - u)$ converge weakly to 0 and that $Q(m_\varepsilon - u) - Q(u_\varepsilon - u) = Q((1 - \rho_\varepsilon)u)$ converges to 0 in $L^2(0, T; H^s)$ ($\forall s \geq 0$), we see that it is equivalent to show the above requirement for the following term $\operatorname{div}(Q(m_\varepsilon - u) \otimes Q(m_\varepsilon - u))$. Next, we introduce ψ_ε such that $\int_B \psi_\varepsilon dx = 0$, $\nabla \psi_\varepsilon = Qm_\varepsilon$. Besides, it is easy to see that ψ_ε is bounded in $L^\infty(0, T; H^s)$ ($\forall s \geq 0$). With the above notations, we deduce from the initial system (69) the following one

$$\frac{\partial \varphi_\varepsilon}{\partial t} + \frac{1}{\varepsilon} \Delta \psi_\varepsilon = 0 \quad \frac{\partial \nabla \psi_\varepsilon}{\partial t} + \frac{a\gamma}{\varepsilon} \nabla \varphi_\varepsilon = F_\varepsilon, \quad (100)$$

where $F_\varepsilon = \xi \nabla \operatorname{div} u_\varepsilon + \nabla \pi_\varepsilon + \mu Q[\Delta u_\varepsilon - \operatorname{div}(\rho_\varepsilon u_\varepsilon \otimes u_\varepsilon)]$ is bounded in $L^2(0, T; H^s)$ ($\forall s \geq 0$).

Next, we observe that in $\mathcal{D}'(B \times]0, T[)$, we have on one hand,

$$\operatorname{div}(Qu \otimes Qu) = \frac{1}{2} \nabla |Qu|^2 + (\operatorname{div} Qu) Qu = \frac{1}{2} \nabla |Qu|^2$$

and on the other hand,

$$\begin{aligned} \operatorname{div}(\nabla \psi_\varepsilon \otimes \nabla \psi_\varepsilon) &= \frac{1}{2} \nabla |\nabla \psi_\varepsilon|^2 + \Delta \psi_\varepsilon \nabla \psi_\varepsilon \\ &= \frac{1}{2} \nabla (|\nabla \psi_\varepsilon|^2) - \frac{\partial}{\partial t} (\varepsilon \varphi_\varepsilon \nabla \psi_\varepsilon) + \varepsilon \varphi_\varepsilon F_\varepsilon - a\gamma \varphi_\varepsilon \nabla \varphi_\varepsilon \\ &= \frac{1}{2} \nabla (|\nabla \psi_\varepsilon|^2 - a\gamma \varphi_\varepsilon^2) - \frac{\partial}{\partial t} (\varepsilon \varphi_\varepsilon \nabla \psi_\varepsilon) + \varepsilon \varphi_\varepsilon F_\varepsilon. \end{aligned}$$

Using that $\varepsilon\varphi_\varepsilon\nabla\psi_\varepsilon$ converges strongly to 0 in $L^2(0, T; H^s)$ ($\forall s \geq 0$) and that $\varepsilon\varphi_\varepsilon F_\varepsilon$ converges strongly to 0 in $L^\infty(0, T; H^s)$ ($\forall s \geq 0$), we deduce that

$$\operatorname{div}(\varrho(m_\varepsilon - u) \otimes \varrho(m_\varepsilon - u)) \xrightarrow{\varepsilon} \nabla q \quad (101)$$

and finally, we obtain that

$$\operatorname{div}(\rho_\varepsilon u_\varepsilon \otimes u_\varepsilon) \xrightarrow{\varepsilon} \operatorname{div}(u \otimes u) + \nabla q \quad \text{in } B \times (0, T) \quad (102)$$

and the theorem is proved in the regular case. We only notice here that if Ω is not simply connected, we can take C an annulus around each hole in the previous argument to make sure that the pressure is globally well defined.

Convergence in the general case. Now, we are going to show how we can regularize in x the above quantities (uniformly in ε). To do so let $K_\delta = \frac{1}{\delta^N} K(\cdot/\delta)$, where $K \in C_0^\infty(\mathbb{R}^N)$, $\int_{\mathbb{R}^N} K \, dz = 1$, $\delta \in (0, 1)$. We can then regularize by convolution as follows $\varphi_\varepsilon^\delta = \varphi_\varepsilon * K_\delta$, $m_\varepsilon^\delta = m_\varepsilon * K_\delta$, $u_\varepsilon^\delta = u_\varepsilon * K_\delta$, $\pi_\varepsilon^\delta = \pi_\varepsilon * K_\delta$. We can then follow the same proof as in the regular case by replacing φ_ε , π_ε , m_ε and u_ε by their regularizations and we conclude by observing that $\|u_\varepsilon^\delta - u_\varepsilon\|_{L^2(0, T; L^2)} \leq C\delta$, $\|u_\varepsilon^\delta\|_{L^2(0, T; H^1)} \leq C$ and $\|u_\varepsilon^\delta \otimes u_\varepsilon^\delta - u_\varepsilon \otimes u_\varepsilon\|_{L^1(0, T; L^p)} \leq C\delta$, $\|u_\varepsilon^\delta \otimes u_\varepsilon^\delta\|_{L^1(0, T; L^p)} \leq C$ ($p = N/(N-2)$ if $N \geq 3$, $1 \leq p < +\infty$ if $N = 2$). Indeed, from the above uniform bounds, we deduce that

$$\begin{aligned} & \sup_{\varepsilon \in [0, 1]} \{ \|\rho_\varepsilon^\delta u_\varepsilon^\delta - m_\varepsilon^\delta\|_{L^2(L^q)} + \|m_\varepsilon^\delta - u_\varepsilon^\delta\|_{L^2(L^q)} + \|\rho_\varepsilon u_\varepsilon - m_\varepsilon^\delta\|_{L^2(L^q)} \} \xrightarrow{\delta} 0, \\ & \sup_{\varepsilon \in [0, 1]} \{ \|\rho_\varepsilon^\delta u_\varepsilon^\delta \otimes u_\varepsilon^\delta - m_\varepsilon^\delta \otimes u_\varepsilon^\delta\|_{L^1(L^r)} + \|m_\varepsilon^\delta \otimes u_\varepsilon^\delta - m_\varepsilon \otimes u_\varepsilon\|_{L^1(L^r)} \\ & \quad + \|m_\varepsilon \otimes u_\varepsilon - u_\varepsilon^\delta \otimes u_\varepsilon^\delta\|_{L^1(L^r)} \} \xrightarrow{\delta} 0, \end{aligned}$$

with $1/q > 1/\gamma + (N-2)/(2N)$, $1/r > 1/\gamma + (N-2)/N$, since $1/\gamma + (N-2)/N < 1$. Moreover, it is easy to see that for all δ and all s , we have that $\|m_\varepsilon^\delta - u_\varepsilon^\delta\|_{L^2(H^s)}$ goes to 0 when ε goes to 0.

In the next three subsections, we would like to specify the boundary conditions and give a more precise convergence result. $\square$

3.3.2. The periodic case. The periodic case was treated in [111]. The convergence stated in Theorem 3.7 cannot be improved. Indeed, the acoustic waves will oscillate indefinitely. So, we only have weak convergence. The initial condition in (97) can be specified precisely, namely $u^0 = P\tilde{u}^0$.

3.3.3. The case of Dirichlet boundary conditions. In this subsection we will state more precise results in the case of Dirichlet boundary conditions. Indeed, depending on some geometrical property of the domain, we can prove a strong convergence result toward the incompressible Navier–Stokes system, which means that all the oscillations are damped in

the limit. Let Ω be a bounded domain. We consider the system (69) with the following Dirichlet boundary condition

$$u_\varepsilon = 0 \quad \text{on } \partial\Omega. \quad (103)$$

For $\varepsilon \in (0, 1]$, we consider $(\rho_\varepsilon, u_\varepsilon)$ satisfying the same hypotheses as in Section 3.3.1. In order to state precisely our main theorem, we need to introduce a geometrical condition on Ω . Let us consider the following over determined problem

$$-\Delta\phi = \lambda\phi \quad \text{in } \Omega, \quad \frac{\partial\phi}{\partial\mathbf{n}} = 0 \quad \text{on } \partial\Omega, \quad \text{and} \quad \phi \text{ is constant on } \partial\Omega. \quad (104)$$

A solution of (104) is said to be trivial if $\lambda = 0$ and ϕ is a constant. We will say that Ω satisfies such an assumption (H) if all the solutions of (104) are trivial. Schiffer's conjecture says that every Ω satisfies (H) excepted the ball (see for instance [71]). In two-dimensional space, it is proved that every bounded, simply connected open set $\Omega \subset \mathbb{R}^2$ whose boundary is Lipschitz but not real analytic satisfies (H), hence property (H) is generic in $\mathbb{R}^2$. The main result reads as follows.

THEOREM 3.9. *Under the above conditions, ρ_ε converges to 1 in $C([0, T]; L^Y(\Omega))$ and extracting a subsequence if necessary u_ε converges weakly to u in $L^2((0, T) \times \Omega)^N$ for all $T > 0$, and strongly if Ω satisfies (H). In addition, u is a global weak solution of the incompressible Navier–Stokes equations with Dirichlet boundary conditions satisfying $u|_{t=0} = P\tilde{u}_0$ in Ω .*

For the proof of this result, we refer to [54]. We only sketch below the phenomenon going on. Let $(\lambda_{k,0}^2)_{k \geq 1}$, $\lambda_{k,0} > 0$, be the nondecreasing sequence of eigenvalues and $(\Psi_{k,0})_{k \geq 1}$ the orthonormal basis of $L^2(\Omega)$ functions with zero mean value of eigenvectors of the Laplace operator $-\Delta_N$ with homogeneous Neumann boundary conditions

$$-\Delta\Psi_{k,0} = \lambda_{k,0}^2\Psi_{k,0} \quad \text{in } \Omega, \quad \partial\Psi_{k,0}\partial\mathbf{n} = 0 \quad \text{on } \partial\Omega. \quad (105)$$

We can split these eigenvectors $(\Psi_{k,0})_{k \in \mathbb{N}}$ (which represent the acoustic eigenmodes in Ω) into two classes: those which are not constant on $\partial\Omega$ will generate boundary layers and will be quickly damped, thus converging strongly to 0; those which are constant on $\partial\Omega$ (nontrivial solutions of (104)), for which no boundary layer forms, will remain oscillating forever, leading to only weak convergence. Indeed, if (H) is not satisfied, u^ε will in general only converge weakly and not strongly to u (like in the periodic case $\Omega = \mathbb{T}^d$ for instance). However, if at initial time $t = 0$, no modes of second type are present in the velocity, the convergence to the incompressible solution is strong in L^2 .

Notice that according to Schiffer's conjecture the convergence is not strong for general initial data when Ω is the two- or three-dimensional ball, but is expected to be always strong in any other domain with Dirichlet boundary conditions.

3.3.4. The whole space case. In [52], the authors give a more precise result in the whole space case by using the dispersion of the acoustic waves.

Consider the system (95) in the whole space $\mathbb{R}^N$. The initial data $(\rho_\varepsilon^0, m_\varepsilon^0)$ satisfies

$$\int_{\mathbb{R}^N} \pi_\varepsilon(t=0) + \frac{|m_\varepsilon^0|^2}{2\rho_\varepsilon^0} dx \leq C, \quad (106)$$

where $\pi_\varepsilon = \frac{1}{\gamma(\gamma-1)\varepsilon^2}(\rho_\varepsilon^\gamma - 1 - \gamma(\rho_\varepsilon - 1))$, $m_\varepsilon = \rho_\varepsilon u_\varepsilon$. We also assume that $m_\varepsilon^0/\sqrt{\rho_\varepsilon^0}$ converges weakly in $L^2(\mathbb{R}^N)$ to some $\tilde{u}^0$. Let $L_2^p(\mathbb{R}^N)$ denote the Orlicz space $L_2^p(\mathbb{R}^N) = \{f \in L_{\text{loc}}^1(\mathbb{R}^N)/f\mathbb{1}_{|f|<1} \in L^2 \text{ and } f\mathbb{1}_{|f|\geq 1} \in L^p\}$. We consider global weak solutions to (95) with the initial data (106) satisfying (96) with Ω replaced by $\mathbb{R}^N$ and such that $\rho_\varepsilon - 1 \in L^\infty(0, T; L_2^\gamma(\mathbb{R}^N))$.

THEOREM 3.10. *Under the above assumptions, $\rho_\varepsilon - 1$ converges to 0 in $L^\infty(0, T; L_2^\gamma)$. For all subsequence of u_ε which converges weakly to some $u \in L^2$, u is a global weak solution of the incompressible Navier–Stokes system with the initial data $u(t=0) = P\tilde{u}^0$. Moreover, the subsequence u_ε converges strongly to u in $L^2(0, T; L^2(\mathbb{R}_{\text{loc}}^N))$ and the gradient part Qu_ε converges strongly to 0 in $L^2(0, T; L^q(\mathbb{R}^N))$ for $q > 2$ when $N = 2$ and for $q \in (2, 6)$ when $N = 3$.*

The proof uses the Strichartz estimate (84) to prove that the acoustic waves locally go to zero.

3.3.5. Convergence toward the Euler system. In this subsection, we study the case where μ_ε goes to 0 too. We will state two results in the periodic case and in the whole space case taken from [127]. The case of domains with boundaries is open even in the incompressible case (see Section 2).

The whole space case. We consider a sequence of global weak solutions $(\rho_\varepsilon, u_\varepsilon)$ of the compressible Navier–Stokes equations (69) and we assume that $\rho_\varepsilon - 1 \in L^\infty(0, \infty; L_2^\gamma) \cap C([0, \infty), L_2^p)$ for all $1 \leq p < \gamma$, where $L_2^p = \{f \in L_{\text{loc}}^1, |f|\mathbb{1}_{|f|\geq 1} \in L^p, |f|\mathbb{1}_{|f|\leq 1} \in L^2\}$, $u_\varepsilon \in L^2(0, T; H^1)$ for all $T \in (0, \infty)$ (with a norm which can explode when ε goes to 0), $\rho_\varepsilon |u_\varepsilon|^2 \in L^\infty(0, \infty; L^1)$ and $\rho_\varepsilon u_\varepsilon \in C([0, \infty); L^{2\gamma/(\gamma+1)} - w)$, i.e., is continuous with respect to $t \geq 0$ with values in $L^{2\gamma/(\gamma+1)}$ endowed with its weak topology. We require (69) to hold in the sense of distributions and we impose the following conditions at infinity

$$\rho_\varepsilon \rightarrow 1 \quad \text{as } |x| \rightarrow +\infty, \quad u_\varepsilon \rightarrow 0 \quad \text{as } |x| \rightarrow +\infty. \quad (107)$$

Finally, we prescribe initial conditions $\rho_\varepsilon(t=0) = \rho_\varepsilon^0$, $\rho_\varepsilon u_\varepsilon(t=0) = m_\varepsilon^0$, where $\rho_\varepsilon^0 \geq 0$, $\rho_\varepsilon^0 - 1 \in L^\gamma$, $m_\varepsilon^0 \in L^{2\gamma/(\gamma+1)}$, $m_\varepsilon^0 = 0$ a.e. on $\{\rho_\varepsilon^0 = 0\}$ and $\rho_\varepsilon^0 |u_\varepsilon^0|^2 \in L^1$, denoting by $u_\varepsilon^0 = m_\varepsilon^0/\rho_\varepsilon^0$ on $\{\rho_\varepsilon^0 > 0\}$, $u_\varepsilon^0 = 0$ on $\{\rho_\varepsilon^0 = 0\}$. We also introduce the following notation

$\rho_\varepsilon = 1 + \varepsilon\varphi_\varepsilon$. Notice that if $\gamma < 2$, we cannot deduce any bound for φ_ε in $L^\infty(0, T; L^2)$. This is why we introduce the following approximation which belongs to L^2

$$\Phi_\varepsilon = \frac{1}{\varepsilon} \sqrt{\frac{2a}{\gamma-1} (\rho_\varepsilon^\gamma - 1 - \gamma(\rho_\varepsilon - 1))}.$$

Furthermore, we assume that $\sqrt{\rho_\varepsilon^0} u_\varepsilon^0$ converges strongly in L^2 to some $\tilde{u}^0$. Then, we denote by $u^0 = P\tilde{u}^0$, where P is the projection on divergence-free vector fields, we also define Q (the projection on gradient vector fields), hence $\tilde{u}^0 = P\tilde{u}^0 + Q\tilde{u}^0$. Moreover, we assume that Φ_ε^0 converges strongly in L^2 to some φ^0 . This also implies that φ_ε^0 converges to φ^0 in L_2^γ . We also assume that $(\rho_\varepsilon, u_\varepsilon)$ satisfies the energy inequality. Our last requirement on $(\rho_\varepsilon, u_\varepsilon)$ concerns the total energy: we assume that we have

$$E_\varepsilon(t) + \int_0^t D_\varepsilon(s) ds \leq E_\varepsilon^0 \quad \text{a.e. } t, \quad \frac{dE_\varepsilon}{dt} + D_\varepsilon \leq 0 \quad \text{in } \mathcal{D}'(0, \infty), \quad (108)$$

where

$$E_\varepsilon(t) = \int_\Omega \frac{1}{2} \rho_\varepsilon |u_\varepsilon|^2(t) + \frac{a}{\varepsilon^2(\gamma-1)} ((\rho_\varepsilon)^\gamma - 1 - \gamma(\rho_\varepsilon - 1))(t),$$

$$D_\varepsilon(t) = \int_\Omega \mu_\varepsilon |Du_\varepsilon|^2(t) + \xi_\varepsilon (\operatorname{div} u_\varepsilon)^2(t)$$

and

$$E_\varepsilon^0 = \int_\Omega \frac{1}{2} \rho_\varepsilon^0 |u_\varepsilon^0|^2 + \frac{a}{\varepsilon^2(\gamma-1)} ((\rho_\varepsilon^0)^\gamma - 1 - \gamma(\rho_\varepsilon^0 - 1)).$$

The existence of solutions satisfying the above requirement was proved in [109].

When ε goes to zero and μ_ε goes to 0, we expect that u_ε converges to v , the solution of the Euler system

$$\begin{cases} \partial_t v + \operatorname{div}(v \otimes v) + \nabla \pi = 0, \\ \operatorname{div} v = 0, \quad v|_{t=0} = u^0, \end{cases} \quad (109)$$

in $C([0, T^*]; H^s)$. We have the following theorem.

THEOREM 3.11. *We assume that $\mu_\varepsilon \xrightarrow{\varepsilon} 0$ (such that $\mu_\varepsilon + \xi_\varepsilon > 0$ for all ε) and that $P\tilde{u}^0 \in H^s$ for some $s > N/2 + 1$, then $P(\sqrt{\rho_\varepsilon} u_\varepsilon)$ converges to v in $L^\infty(0, T; L^2)$ for all $T < T^*$, where v is the unique solution of the Euler system in $L_{\text{loc}}^\infty([0, T^*]; H^s)$ and T^* is the existence time of (109). In addition $\sqrt{\rho_\varepsilon} u_\varepsilon$ converges to v in $L^p(0, T; L_{\text{loc}}^2)$ for all $1 \leq p < +\infty$ and all $T < T^*$.*

The periodic case. Now, we take $\Omega = \mathbb{T}^N$ and consider a sequence of solutions $(\rho_\varepsilon, u_\varepsilon)$ of (69), satisfying the same conditions as in the whole space case (the functions are now periodic in space and all the integration are performed over $\mathbb{T}^N$). Of course, the conditions at infinity are removed and the spaces L^p_2 can be replaced by L^p . Here, we have to impose more conditions on the oscillating part (acoustic waves), namely we have to assume that $Q\tilde{u}^0$ is more regular than L^2 . In fact, in the periodic case, we do not have a dispersion phenomenon as in the case of the whole space and the acoustic waves will not go to infinity, but they are going to interact with each other. This is why, we have to include them in the energy estimates to show our convergence result. This requires an analysis of the possible resonances between the different modes.

For the next theorem, we assume that $Q\tilde{u}^0, \varphi^0 \in H^{s-1}$ and that there exists a nonnegative constant ν such that $\mu_\varepsilon + \xi_\varepsilon \geq 2\nu > 0$ for all ε . For simplicity, we assume that $\mu_\varepsilon + \xi_\varepsilon$ converges to 2ν .

THEOREM 3.12 (The periodic case). *We assume that $\mu_\varepsilon \xrightarrow{\varepsilon} 0$ (such that $\mu_\varepsilon + \xi_\varepsilon \rightarrow 2\nu > 0$) and that $P\tilde{u}^0 \in H^s$ for some $s > N/2 + 1$, and $Q\tilde{u}^0, \varphi^0 \in H^{s-1}$ then $P(\sqrt{\rho_\varepsilon}u_\varepsilon)$ converges to v in $L^\infty(0, T; L^2)$ for all $T < T^*$, where v is the unique solution of the Euler system in $L^\infty_{\text{loc}}(0, T^*; H^s)$ and T^* is the existence time of (109). In addition, $\sqrt{\rho_\varepsilon}u_\varepsilon$ converges weakly to v in $L^\infty(0, T; L^2)$.*

IDEA OF THE PROOFS. The proofs of Theorems 3.11 and 3.12 are based on energy estimates, since we loose the compactness in x from the viscosity at the limit. Indeed, using the energy bounds, we deduce that $\rho_\varepsilon - 1$ converges to 0 in $L^\infty(0, T; L^2_\gamma)$ and that there exists some $u \in L^\infty(0, T; L^2)$ and a subsequence $\sqrt{\rho_\varepsilon}u_\varepsilon$ converging weakly to u . Hence, we also deduce that $\rho_\varepsilon u_\varepsilon$ converges weakly to u in $L^{2\gamma/(\gamma+1)}$. Here we are in a situation where we do not have compactness in time and we do not have compactness in space. This is why we have to use an energy method. For this, we have to describe the oscillations in time and incorporate them in the energy estimates. It turns out that in the whole space case the acoustic waves disperse to infinity as can be deduced from the Strichartz estimate (84). We also refer to [166] and Theorem 3.3 in the framework of strong solutions and [52] and Theorem 3.10 in the framework of weak solutions. In the sequel, we will concentrate more on the periodic case. The operators L and $\mathcal{L}$ were defined in (83). Let

$$U^\varepsilon = (\varphi_\varepsilon, Q(\rho_\varepsilon u_\varepsilon)) \quad \text{and} \quad V^\varepsilon = \mathcal{L}\left(-\frac{t}{\varepsilon}\right)(\varphi_\varepsilon, Q(\rho_\varepsilon u_\varepsilon)).$$

Using that

$$\varepsilon \frac{\partial \varphi_\varepsilon}{\partial t} + \operatorname{div} Q(\rho_\varepsilon u_\varepsilon) = 0, \quad \varepsilon \frac{\partial}{\partial t} Q(\rho_\varepsilon u_\varepsilon) + \nabla \varphi_\varepsilon = \varepsilon F_\varepsilon \quad (110)$$

for some F_ε which is bounded in $L^2 H^{-r}$ for some $r \in \mathbb{R}$, we deduce that $\partial_t U^\varepsilon = \frac{1}{\varepsilon} L U^\varepsilon + (0, F_\varepsilon)$, and hence that $\partial_t V^\varepsilon = \mathcal{L}(-t/\varepsilon)(0, F_\varepsilon)$. This means that V^ε is compact in time since the oscillations have been canceled by $\mathcal{L}(-t/\varepsilon)$. If we had enough compactness in

space we could pass to the limit in this equation and recover the following limit system for the oscillating part

$$\partial_t \bar{V} + \mathcal{Q}_1(u, \bar{V}) + \mathcal{Q}_2(\bar{V}, \bar{V}) - \nu \Delta \bar{V} = 0, \quad (111)$$

where $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are respectively a linear and a bilinear forms in $\bar{V}$ defined by the following definition. $\square$

DEFINITION 3.13. For all divergence-free vector field $u \in L^2(\Omega)^N$ and all $V = (\psi, \nabla q) \in L^2(\Omega)^{N+1}$, we define the following linear and bilinear symmetric forms in V

$$\begin{aligned} \mathcal{Q}_1(u, V) &= \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau \mathcal{L}(-s) \begin{pmatrix} 0 \\ \operatorname{div}(u \otimes \mathcal{L}_2(s)V + \mathcal{L}_2(s)V \otimes u) \end{pmatrix} ds \end{aligned} \quad (112)$$

and

$$\begin{aligned} \mathcal{Q}_2(V, V) &= \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau \mathcal{L}(-s) \begin{pmatrix} 0 \\ \operatorname{div}(\mathcal{L}_2(s)V \otimes \mathcal{L}_2(s)V) + \frac{\gamma-1}{2} \nabla(\mathcal{L}_1(s)V)^2 \end{pmatrix} ds. \end{aligned} \quad (113)$$

The convergences stated above take place in $W^{-1,1}$ and can be shown by using almost-periodic functions (see [125] and the references therein). We also notice that

$$-\nu \Delta V = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau -\mathcal{L}(-s) \begin{pmatrix} 0 \\ 2\nu \Delta \mathcal{L}_2(s)V \end{pmatrix} ds. \quad (114)$$

To recover compactness in space, we will use the regularity of the limit system. Let V^0 be the solution of the following system

$$\begin{cases} \partial_t V^0 + \mathcal{Q}_1(v, V^0) + \mathcal{Q}_2(V^0, V^0) - \nu \Delta V^0 = 0, \\ V^0|_{t=0} = (\varphi^0, Q\tilde{u}^0), \end{cases} \quad (115)$$

where v is the solution of the incompressible Euler equations with initial data u^0 . The existence of global strong solutions for the system (115) (and local solutions if the viscosity term is removed) can be deduced from the exact computations of the two forms $\mathcal{Q}_1$ and $\mathcal{Q}_2$. We point out that in the case $\nu > 0$, the existence of a global solution to the system (115) is an important property of (115) which is not shared by the Navier–Stokes system from which it is derived. Indeed, the nonlinear term $\mathcal{Q}_2(V^0, V^0)$ can be decomposed into a countable number of Burgers equations. We refer to [127] for more details.

Finally, the energy method is based on the fact that we can apply a Grönwall lemma to the following quantity

$$\left\| \sqrt{\rho_\varepsilon} u_\varepsilon - v - \mathcal{L}_2\left(\frac{t}{\varepsilon}\right)V \right\|_{L^2}^2 + \left\| \Phi_\varepsilon - \mathcal{L}_1\left(\frac{t}{\varepsilon}\right)V \right\|_{L^2}^2. \quad (116)$$

Notice indeed, that from the analysis given above, we expect that $\sqrt{\rho_\varepsilon} u_\varepsilon$ behaves like $v + \mathcal{L}_2(t/\varepsilon)V$ and that ϕ_ε and Φ_ε behave like $\mathcal{L}_1(t/\varepsilon)V$. For the details, we refer to [127].

We want to point out that the method of proof is the same for the whole space case and is simpler since we do not have to study all the resonances (the acoustic waves go to infinity). So, we just need to apply a Grönwall lemma to the quantity given in (116) where V is replaced by $V(t=0)$.

REMARK 3.14. In Theorem 3.12, one can remove the condition $2\nu > 0$. In that case, we still have the result of Theorem 3.12 but only on an interval of time $(0, T^{**})$ which is the existence interval for the equation governing the oscillating part (115). Indeed, it is easy to see using the particular form of $\mathcal{Q}_1$ and $\mathcal{Q}_2$ that if $\nu > 0$ and $V(t=0) \in H^{s-1}$ then, we have (as long as v exists) a global solution in $L^\infty(H^{s-1})$ which satisfies $\nabla V \in L^1(0, T; L^\infty)$. On the other hand, if $\nu = 0$ and $V(t=0) \in H^{s-1}$ then we can only construct a local (in time) solution in $L^\infty(H^{s-1})$ which satisfies $\nabla V \in L^1(0, T; L^\infty)$ for all $T < T^{**}$.

3.4. Study of the limit $\gamma \rightarrow \infty$

In this subsection we are going to study the limit γ going to infinity. Depending on the total mass, we will recover at the limit either a mixed model, which behaves as a compressible one if $\rho < 1$ and as an incompressible one if $\rho = 1$ or the classical incompressible Navier–Stokes system. We start with the first case and define the limit system, namely

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0 \quad \text{in } (0, T) \times \Omega, \quad 0 \leq \rho \leq 1 \quad \text{in } (0, T) \times \Omega, \quad (117)$$

$$\frac{\partial \rho u}{\partial t} + \operatorname{div}(\rho u \otimes u) - \mu \Delta u - \xi \nabla \operatorname{div} u + \nabla \pi = 0 \quad \text{in } (0, T) \times \Omega, \quad (118)$$

$$\operatorname{div} u = 0 \quad \text{a.e. on } \{\rho = 1\}, \quad (119)$$

$$\pi = 0 \quad \text{a.e. on } \{\rho < 1\}, \quad \pi \geq 0 \quad \text{a.e. on } \{\rho = 1\}. \quad (120)$$

In all this section Ω is taken to be the torus, the whole space or a bounded domain with Dirichlet boundary conditions. Indeed, the proofs given in [112] can also apply to the case of Dirichlet boundary conditions, by using the bounds given in [110] and [67].

Let γ_n be a sequence of nonnegative real numbers that goes to infinity. Let (ρ_n, u_n) be a sequence of weak solutions to the isentropic compressible Navier–Stokes equations

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0, & \rho \geq 0, \\ \frac{\partial \rho u}{\partial t} + \operatorname{div}(\rho u \otimes u) - \mu \Delta u - \xi \nabla \operatorname{div} u + \nabla \rho^{\gamma_n} = 0, \end{cases} \quad (121)$$

where $\mu > 0$ and $\mu + \xi > 0$. We recall that global weak solutions of the above system are known to exist, if we assume in addition that $\gamma_n > N/2$. This holds for n large enough. The sequence (ρ_n, u_n) satisfies in addition the following initial conditions and the following bounds,

$$\rho_n u_n(t=0) = m_n^0, \quad \rho_n(t=0) = \rho_n^0, \quad (122)$$

where $0 \leq \rho_n^0$ a.e., ρ_n^0 is bounded in $L^1(\Omega)$ and $\rho_n^0 \in L^{\gamma_n}$ with $\int (\rho_n^0)^{\gamma_n} \leq C\gamma_n$ for some fixed C , $m_n^0 \in L^{2\gamma_n/(\gamma_n+1)}(\Omega)$ and $\rho_n^0 |u_n^0|^2$ is bounded in L^1 , denoting by $u_n^0 = m_n^0/\rho_n^0$ on $\{\rho_n^0 > 0\}$, $u_n^0 = 0$ on $\{\rho_n^0 = 0\}$. In the periodic case or in the Dirichlet boundary condition case, we also assume that $\int \rho_n^0 = M_n$, for some M_n such that $0 < M_n \leq M < 1$ and $M_n \rightarrow M$. Furthermore, we assume that $\rho_n^0 u_n^0$ converges weakly in L^2 to some m^0 and that ρ_n^0 converges weakly in L^1 to some ρ^0 . The last requirement concerns the following energy bounds we impose on the sequence of solutions we consider,

$$E_n(t) + \int_0^t D_n(s) ds \leq E_n^0 \quad \text{a.e. } t, \quad \frac{dE_n}{dt} + D_n \leq 0 \quad \text{in } \mathcal{D}'(0, \infty), \quad (123)$$

where

$$E_n(t) = \int \frac{1}{2} \rho_n |u_n|^2(t) + \frac{a}{\gamma_n - 1} (\rho_n)^{\gamma_n}(t),$$

$$D_n(t) = \int \mu |Du_n|^2(t) + \xi (\operatorname{div} u_n)^2(t)$$

and

$$E_n^0 = \int \frac{1}{2} \rho_n^0 |u_n^0|^2 + \frac{a}{\gamma_n - 1} (\rho_n^0)^{\gamma_n}.$$

Without loss of generality, extracting subsequences if necessary, we can assume that (ρ_n, u_n) converges weakly to (ρ, u) . More precisely, we can assume that $\rho_n \rightharpoonup \rho$ weakly in $L^p((0, T) \times \Omega)$ for any $1 \leq p \leq \infty$ and that $\rho \in L^\infty(0, T; L^p)$ (in fact we will show that ρ actually satisfies $0 \leq \rho \leq 1$), $u_n \rightharpoonup u$ weakly in $L^2(0, T; H_{\text{loc}}^1)$.

Before stating the main theorem, we have to define precisely the notion of weak solutions for the limit system. (ρ, u, π) is called a weak solution of the limit system (117)–(120) if

$$\rho \in L^\infty(0, T; L^\infty \cap L^1(\Omega)) \cap C(0, T; L^p) \quad \text{for any } 1 \leq p < \infty, \quad (124)$$

$$\nabla u \in L^2(0, T, L^2) \quad \text{and} \quad u \in L^2(0, T; H^1(B)), \quad (125)$$

where $B = \Omega$ if $\Omega = \mathbb{T}^N$ or if Ω is a bounded domain (with Dirichlet boundary conditions) and B is any ball in $\mathbb{R}^N$ if $\Omega = \mathbb{R}^N$, in this last case we also impose that $u \in L^2(0, T, L^{2N/(N-2)}(\mathbb{R}^N))$, if in addition $N \geq 3$. Moreover,

$$\rho |u|^2 \in L^\infty(0, \infty; L^1) \quad \text{and} \quad \rho u \in L^\infty(0, \infty; L^2). \quad (126)$$

Next, equations (117) and (118) must be satisfied in the distributional sense. This can be written using a weak formulation (which also incorporate the initial conditions in some weak sense), namely we require that the following identities hold for all $\phi \in C^\infty([0, \infty) \times \Omega)$ and for all $\Phi \in C^\infty([0, \infty) \times \Omega)^N$ compactly supported in $[0, \infty) \times \Omega$ (i.e., vanishing

identically for t large enough),

$$-\int_0^\infty dt \int_\Omega \rho \partial_t \phi - \int_\Omega \rho^0 \phi(0) - \int_0^\infty dt \int_\Omega \rho u \cdot \nabla \phi = 0, \quad (127)$$

$$\begin{aligned} & -\int_0^\infty dt \int_\Omega \rho u \cdot \partial_t \Phi - \int_\Omega m^0 \cdot \Phi(0) - \int_0^\infty dt \int_\Omega \rho (u \cdot \nabla \Phi) \cdot u \\ & + \int_0^\infty dt \left\{ \int_\Omega \mu Du \cdot D\Phi + \xi \operatorname{div} u \operatorname{div} \Phi \right\} - \pi \operatorname{div} \Phi = 0. \end{aligned} \quad (128)$$

On the other hand, equation (120) should be understood in the following way $\rho\pi = \pi \geq 0$. Of course, we have to define the sense of the product $\rho\pi$ since, we only require that $\pi \in \mathcal{M}$. Indeed, the product can be defined by using that

$$\begin{cases} \rho \in C([0, T]; L^p) \cap C^1([0, T]; H^{-1}), \\ \pi \in W^{-1, \infty}(H^1) + L^1(L^{N/(N-2)}) \cap L^\alpha(L^\beta) + L^2(L^2), \end{cases} \quad (129)$$

where $1 < \alpha, \beta < \infty$ and $1/\beta = 1/\alpha(N-2)/N + (1-1/\alpha)$.

Finally, equation (119) is just a consequence of (117), however we incorporate it in the limit system to emphasize the fact that it is a mixed system which behaves like a compressible one if $\rho < 1$ and as an incompressible one if $\rho = 1$.

THEOREM 3.15. *Under the above conditions, we have $0 \leq \rho \leq 1$ and*

$$(\rho_n - 1)_+ \rightarrow 0 \quad \text{in } L^\infty(0, T; L^p) \text{ for any } 1 \leq p < +\infty.$$

Moreover, $(\rho_n)^{\gamma_n}$ is bounded in L^1 (for n such that $\gamma_n \geq N$). Then extracting subsequences again, there exists $\pi \in \mathcal{M}((0, T) \times \Omega)$ such that

$$(\rho_n)^{\gamma_n} \xrightarrow[n]{\rightharpoonup} \pi. \quad (130)$$

If in addition ρ_n^0 converges in L^1 to ρ^0 then (ρ, u, π) is a weak solution of (117)–(120) and the following strong convergences hold

$$\begin{aligned} \rho_n &\rightarrow \rho && \text{in } C(0, T; L^p(\Omega)) \text{ for any } 1 \leq p < +\infty, \\ \rho_n u_n &\rightarrow \rho u && \text{in } L^p(0, T; L^q(\Omega)) \text{ for any } 1 \leq p < +\infty, \ 1 \leq q < 2, \\ \rho_n u_n \otimes u_n &\rightarrow \rho u \otimes u && \text{in } L^p(0, T; L^1(\Omega)) \text{ for any } 1 \leq p < +\infty. \end{aligned}$$

The second result concerns the case $M > 1$. Let (ρ_n, u_n) be a sequence of solutions of (121) satisfying the above requirement but where we assume now that $\int \rho_n^0 = M > 1$, $\int (\rho_n^0)^{\gamma_n} \leq M^{\gamma_n} + C\gamma_n$ for some fixed C .

THEOREM 3.16. *Under the above assumptions, ρ_n converges to M in $C([0, T]; L^p(\Omega))$ for $1 \leq p < +\infty$, $\sqrt{\rho_n} u_n$ converges weakly to $\sqrt{M}u$ in $L^\infty(0, T; L^2(\Omega))$ and Du_n converges weakly to Du in $L^2(0, T; L^2(\Omega))$ for all $T \in (0, \infty)$ where u is a solution of the incompressible Navier–Stokes system*

$$\begin{aligned} \frac{\partial u}{\partial t} + \operatorname{div}(u \otimes u) - \mu M \Delta u + \nabla p &= 0, \\ \operatorname{div} u &= 0, \quad u|_{t=0} = P(m^0). \end{aligned}$$

For the proof of these two theorems we refer to [112] and to [124] for the Dirichlet boundary condition case.

3.5. The nonisentropic case

We consider the nonisentropic compressible Euler system. This can be written after some simple change of variable in the following form (see [133]):

$$\begin{cases} a(\partial_t q + v \cdot \nabla q) + \frac{1}{\varepsilon} \nabla \cdot v = 0, \\ r(\partial_t v + v \cdot \nabla v) + \frac{1}{\varepsilon} \nabla q = 0, \\ \partial_t S + v \cdot \nabla S = 0, \end{cases} \quad (131)$$

where $a = a(S, \varepsilon q)$ and $r = r(S, \varepsilon q)$ are positive given function of S and εq . In (131), S is the entropy, $P = \underline{P}e^{\varepsilon q}$ is the pressure for some constant $\underline{P}$ and v is a rescaled velocity. The equation of state is given by the density $\rho = R(S, P)$ from which we can deduce the function a and r by

$$a(S, \varepsilon q) = \frac{P}{R} \frac{\partial R(S, P)}{\partial P}, \quad r(S, \varepsilon q) = \frac{R(S, P)}{P}. \quad (132)$$

Formally when ε goes to zero, we expect that the solution $(q_\varepsilon, v_\varepsilon, S_\varepsilon)$ to the system (131) converges to a solution of the following limit system

$$\begin{cases} r_0(S)(\partial_t v + v \cdot \nabla v) + \nabla \pi = 0, \\ \operatorname{div} v = 0, \\ \partial_t S + v \cdot \nabla S = 0, \end{cases} \quad (133)$$

where $r_0(S) = r(S, 0)$. The limit system (133) is an inhomogeneous incompressible Euler system (see [108] for some remarks about this system). This convergence was first proved in the “well-prepared” case in [153].

For general initial data, there are two major questions we can ask about the system (131). Can we solve (131) on some time interval which is independent of ε ? And can we characterize the limit of $(q_\varepsilon, v_\varepsilon, S_\varepsilon)$ when ε goes to zero? For the first question a full satisfactory answer is given in [133]. For the second equation, Métivier and Schochet [133] prove the convergence toward the limit system (133) in the whole space by using the dispersion for a wave equation with non constant coefficients. For the periodic case the problem is much

more involved due to the oscillations in time. In [134] the same authors give some partial results. The case of the exterior domain is treated in [2]. Before stating the result of [133], let us mention the reference [26] where a formal computation is made in the periodic case and the recent paper [3] where the full compressible Navier–Stokes is considered in the whole space.

Let us take some initial data for (131) $(q_\varepsilon, v_\varepsilon, S_\varepsilon)$ $(t = 0) = (q_\varepsilon^0, v_\varepsilon^0, S_\varepsilon^0)$. The following result is proved in [133]. The first part applies to the case $\Omega = \mathbb{T}^N$ and $\Omega = \mathbb{R}^N$ (see also [2] for domains with boundary). The second part is only for the whole space case (see [2] for the case of an exterior domain).

THEOREM 3.17. (i) Assume that $\|(q_\varepsilon^0, v_\varepsilon^0, S_\varepsilon^0)\|_{H^s} \leq M_0$ where $s > N/2 + 1$. There exists $T = T(M_0)$ such that for all $0 < \varepsilon \leq 1$, the Cauchy problem with the initial data $(q_\varepsilon^0, v_\varepsilon^0, S_\varepsilon^0)$ has a unique solution $(q_\varepsilon, v_\varepsilon, S_\varepsilon) \in C([0, T]; H^s)$.

(ii) Moreover if $\Omega = \mathbb{R}^N$ and $(v_\varepsilon^0, S_\varepsilon^0)$ converges in $H^s(\mathbb{R}^N)$ to some (v^0, S^0) and S_ε^0 decays at infinity in the sense

$$|S_\varepsilon^0(x)| \leq C|x|^{-1-\delta}, \quad |\nabla S_\varepsilon^0(x)| \leq C|x|^{-2-\delta},$$

then $(q_\varepsilon, v_\varepsilon, S_\varepsilon)$ converges weakly in $L^\infty(0, T; H^s)$ and strongly in $L^2(0, T; H_{\text{loc}}^{s'})$ for all $s' < s$ to a limit $(0, v, S)$. Moreover, (v, S) is the unique solution in $C([0, T]; H^s)$ of the limit system (133) with the initial data (w_0, S_0) where w_0 is the unique solution in $H^s(\mathbb{R}^N)$ of

$$\operatorname{div}(w_0) = 0, \quad \operatorname{curl}(r_0 w_0) = \operatorname{curl}(r_0 v_0), \quad \text{where } r_0 = r(S_0, 0). \quad (134)$$

The difficulty in proving the convergence toward the limit system is that the acoustic waves satisfy a wave equation with variable coefficients. The proof of the convergence is based on the use of the H -measures (which were introduced by Gérard [72] and Tartar [161]) to analysis the oscillating part and actually prove that it disperses to infinity as was the case in the isentropic case.

4. Study of rotating fluids at high frequency

In this section we will study rotating fluids when the frequency of rotation goes to zero. This is a singular limit which has many similarities with the compressible–incompressible limit. We will not detail all the known results for this system. We consider the following system of equations

$$\partial_t u^n + \operatorname{div}(u^n \otimes u^n) - \nu \partial_z^2 u^n - \eta \Delta_{x,y} u^n + \frac{e_3 \times u^n}{\varepsilon} = -\frac{\nabla p}{\varepsilon} + F \quad \text{in } \Omega, \quad (135)$$

$$\operatorname{div}(u^n) = 0 \quad \text{in } \Omega, \quad (136)$$

$$u^n(0) = u_0^n \quad \text{with } \operatorname{div}(u_0^n) = 0, \quad (137)$$

$$u^n = 0 \quad \text{on } \partial\Omega, \quad (138)$$

where for example $\Omega = \mathbb{T}^2 \times]0, h[$ or $\Omega = \mathbb{T}^3$, $\nu = \nu^n$ and $\eta = \eta^n$ are respectively the vertical and horizontal viscosities and $\varepsilon = \varepsilon^n$ is the Rossby number. This system describes the motion of a rotating fluid as the Ekman and Rossby numbers go to zero (see [144] and [80]). It can model the ocean, the atmosphere, or a rotating fluid in a container. As for the compressible–incompressible limit the limit system can depend on the boundary conditions in a nontrivial way.

4.1. The periodic case

When there is no boundary ($\Omega = \mathbb{T}^3$ for instance) and when $\nu = \eta = 1$ (the Navier–Stokes case) or $\nu = \eta = 0$ (the Euler case), the problem was studied by several authors ([8–10, 33, 63, 68, 81, 143] ...) by using the group method of [154] and [81]. This method was first introduced to treat the compressible incompressible limit (see Sections 3.2.1 and 3.3.5). Basically, denoting $Lu = -P(e_3 \times u)$ and $\mathcal{L}(\tau) = e^{\tau L}$, we see that $v^n = \mathcal{L}(-t/\varepsilon)u^n$ satisfies

$$\partial_t v^n + \mathcal{L}\left(-\frac{t}{\varepsilon}\right)\left[\operatorname{div}(u^n \otimes u^n) - \nu \partial_z^2 u^n - \eta \Delta_{x,y} u^n\right] = -\nabla q \quad \text{in } \Omega \quad (139)$$

which gives compactness in time for v^n .

The special structure of the limit system which is similar to (159) allows to prove results about long time existence for the Navier–Stokes system when ε goes to zero. This means in some sense that the rotation has a regularizing effect. This regularizing effect also appear when we deal with boundary layers (see the next subsection).

The method introduced in [154] fails when Ω has a boundary (except in very particular cases where there is no boundary layer, or where boundary layers can be eliminated by symmetry [23]).

4.2. Ekman boundary layers in $\Omega = \mathbb{T}^2 \times]0, h[$

In domains with boundaries (for instance $\Omega = \mathbb{T}^2 \times]0, h[$), the case of “well-prepared” initial data was treated in [37, 73, 84, 123]. Here “well-prepared” initial data means that $Lu_0 = 0$ which implies that the initial data is bidimensional and only depends on the horizontal variables. Notice that this implies that there are no oscillations in time. In this case a boundary layer appears at $z = 0$ and $z = h$ to match the nonslip boundary condition with the interior flow. This boundary layer is responsible of the so-called Ekman damping. Let us give a formal expansion leading to the Ekman boundary layer in the well-prepared case (see [84]).

4.2.1. Formal expansion. For convenience we will take here $\varepsilon = \nu$, otherwise there is not such a formal development. Let us write u^n , p and F in the following form

$$U = U^0\left(t, x, y, z, \frac{z}{l}, \frac{h-z}{l}\right) + \varepsilon U^1 + \dots,$$

where l is the length of the boundary layer. Notice here that we do not have a dependence on t/ε since we are concerned here with the well-prepared case. U^0 is decomposed as

$$U^0 = \underline{U}^0(t, x, y, z) + \tilde{U}^0(t, x, y, \theta) + \check{U}^0(t, x, y, \lambda)$$

is the sum of an interior term $\underline{U}^0$ and of two boundary layer terms $\tilde{U}^0$ and $\check{U}^0$ respectively near $z = 0$ and $z = h$, where we set $\theta = z/l$ and $\lambda = (h - z)/l$. We enforce

$$\lim_{\theta \rightarrow \infty} \tilde{U} = 0 \quad \text{and} \quad \lim_{\lambda \rightarrow \infty} \check{U} = 0$$

and, to get the good limit conditions at $z = 0$ and $z = h$,

$$\underline{u}^0(t, x, y, z = 0) + \tilde{u}^0(t, x, y, \theta = 0) = 0, \quad (140)$$

$$\underline{u}^0(t, x, y, z = h) + \check{u}^0(t, x, y, \lambda = 0) = 0. \quad (141)$$

Since the Ekman boundary layers come from the interaction between the viscosity $\nu \partial_z^2 u$ and the Coriolis force $\varepsilon^{-1}(e_3 \times u)$, we take $l = \sqrt{\varepsilon \nu}$, hence $l = \varepsilon$, in this section. Let us focus on the boundary layer near $z = 0$. At the leading order ε^{-2} , one gets

$$\partial_\theta \tilde{p}^0 = 0 \quad \text{hence} \quad \tilde{p}^0 = 0.$$

The pressure does not change in the boundary layer, which is classical in Fluid Mechanics. One also has from (135),

$$-\underline{u}_2^0 = -\partial_x \underline{p}^0, \quad (142)$$

$$\underline{u}_1^0 = -\partial_y \underline{p}^0, \quad (143)$$

$$0 = -\partial_z \underline{p}^0, \quad (144)$$

$$-\partial_\theta^2 \tilde{u}_1^0 - \tilde{u}_2^0 = 0, \quad (145)$$

$$-\partial_\theta^2 \tilde{u}_2^0 + \tilde{u}_1^0 = 0, \quad (146)$$

$$-\partial_\theta^2 \tilde{u}_3^0 = -\partial_\theta \tilde{p}^1 \quad (147)$$

and from (136),

$$\partial_\theta \tilde{u}_3^0 = 0 \quad \text{hence} \quad \tilde{u}_3^0 = 0, \quad (148)$$

$$\partial_x \underline{u}_1^0 + \partial_y \underline{u}_2^0 + \partial_z \underline{u}_3^0 = 0, \quad (149)$$

$$\partial_x \tilde{u}_1^0 + \partial_y \tilde{u}_2^0 + \partial_\theta \tilde{u}_3^1 = 0. \quad (150)$$

Then we obtain from (144) that $\underline{p}^0$ does not depend on z , and from (147) and (148) that $\partial_\theta \tilde{p}^1 = 0$ and hence that $\tilde{p}^1 = 0$. Therefore (142) and (143) give that $\underline{u}_1^0$ and $\underline{u}_2^0$ do not

depend on z , and that

$$\partial_x \underline{u}_1^0 + \partial_y \underline{u}_2^0 = 0. \quad (151)$$

Subtracting this from (149), one gets that $\underline{u}_3^0$ does not depend on z , and since $\tilde{u}_3^0 = 0$, (140) leads to $\underline{u}_3^0 = 0$.

Hence $\underline{u}^0$ satisfies an equation of 2D Navier–Stokes' type. To find this equation, one must take the next order of (42), which gives

$$\partial_t \underline{u}^0 + \nabla(\underline{u}^0 \otimes \underline{u}^0) - \Delta_{x,y} \underline{u}^0 + \begin{pmatrix} -\underline{u}_2^1 \\ \underline{u}_1^1 \\ 0 \end{pmatrix} = -\nabla p^1 + F^0(t, x, y, z) \quad \text{in } \omega. \quad (152)$$

We will suppose that F^0 does not depend on z and that $F_3^0(t, x, y) = 0$. The third component gives that p^1 does not depend on z . Combining this with (152), one finds that $\underline{u}_1^1$ and $\underline{u}_2^1$ do not depend on z . Hence the divergence-free condition for $\underline{u}^1$ shows that $\underline{u}_3^1$ is affine.

Let $\zeta^0 = \text{curl } \underline{u}^0$. We have

$$\partial_t \zeta^0 + (\underline{u}^0 \cdot \nabla) \zeta^0 - \Delta_{x,y} \zeta^0 - \text{curl } F^0 = -\partial_x u_1^1 - \partial_y u_2^1 = \partial_z u_3^1.$$

Integrating this equation with respect to z , we obtain

$$\partial_t \zeta^0 + (\underline{u}^0 \cdot \nabla) \zeta^0 - \Delta_{x,y} \zeta^0 - \text{curl } F^0 = h^{-1} (u_3^1(z=h) - u_3^1(z=0)).$$

Therefore there is a source term in the equation of the vorticity, term which is given by the vertical velocity of the fluid just outside the Ekman boundary layer. So let us compute the boundary layer $\tilde{u}^0$, which satisfies

$$\begin{cases} \partial_\theta^2 \tilde{u}_1^0 = -\tilde{u}_2^0, \\ \partial_\theta^2 \tilde{u}_2^0 = +\tilde{u}_1^0, \\ \tilde{u}_1^0(\theta=0) = -\underline{u}_1^0, & \lim_{\theta \rightarrow \infty} \tilde{u}_1^0 = 0, \\ \tilde{u}_2^0(\theta=0) = -\underline{u}_2^0, & \lim_{\theta \rightarrow \infty} \tilde{u}_2^0 = 0. \end{cases}$$

The solution is given by

$$\begin{cases} \tilde{u}_1^0 = -e^{-\theta/\sqrt{2}} (\underline{u}_1^0 \cos \frac{\theta}{\sqrt{2}} + \underline{u}_2^0 \sin \frac{\theta}{\sqrt{2}}), \\ \tilde{u}_2^0 = -e^{-\theta/\sqrt{2}} (\underline{u}_2^0 \cos \frac{\theta}{\sqrt{2}} - \underline{u}_1^0 \sin \frac{\theta}{\sqrt{2}}). \end{cases}$$

Reporting this in (150) and using (151), one gets

$$\partial_\theta \tilde{u}_3^1 = e^{-\theta/\sqrt{2}} (\partial_x \underline{u}_2^0 - \partial_y \underline{u}_1^0) \sin \frac{\theta}{\sqrt{2}}.$$

Integrating this equation,

$$\tilde{u}_3^1 = -\frac{e^{-\theta/\sqrt{2}}}{\sqrt{2}}(\partial_x \underline{u}_2^0 - \partial_y \underline{u}_1^0) \left(\sin \frac{\theta}{\sqrt{2}} + \cos \frac{\theta}{\sqrt{2}} \right). \quad (153)$$

The integration constant is 0, because $\lim_{\theta \rightarrow \infty} \tilde{u}_3^0 = 0$.

The same calculus holds for the boundary layer at $z = h$, if we change θ by λ and ∂_θ by $-\partial_\lambda$,

$$\begin{cases} \check{u}_1^0 = -e^{-\lambda/\sqrt{2}}(\underline{u}_1^0 \cos \frac{\lambda}{\sqrt{2}} + \underline{u}_2^0 \sin \frac{\lambda}{\sqrt{2}}), \\ \check{u}_2^0 = -e^{-\lambda/\sqrt{2}}(\underline{u}_2^0 \cos \frac{\lambda}{\sqrt{2}} - \underline{u}_1^0 \sin \frac{\lambda}{\sqrt{2}}), \\ \check{u}_3^1 = \frac{e^{-\lambda/\sqrt{2}}}{\sqrt{2}}(\partial_x \underline{u}_2^0 - \partial_y \underline{u}_1^0) \left(\sin \frac{\lambda}{\sqrt{2}} + \cos \frac{\lambda}{\sqrt{2}} \right). \end{cases}$$

Using the limit conditions, and the fact that $\underline{u}_3^1$ is affine, one gets

$$\underline{u}_3^1 = \frac{(\partial_x \underline{u}_2^0 - \partial_y \underline{u}_1^0)}{\sqrt{2}} \left(1 - \frac{2z}{h} \right), \quad (154)$$

$$\partial_z \underline{u}_3^1 = -\frac{\sqrt{2}}{h}(\partial_x \underline{u}_2^0 - \partial_y \underline{u}_1^0). \quad (155)$$

Coming back to (152), we find the limit system

$$\partial_t \underline{u}^0 + \nabla(\underline{u}^0 \otimes \underline{u}^0) - \Delta_{x,y} \underline{u}^0 + \frac{\sqrt{2}}{h} \underline{u}^0 = -\nabla q + F^0 \quad \text{in } \omega. \quad (156)$$

Hence $(\underline{u}_1^0, \underline{u}_2^0)$ satisfies a 2D Navier–Stokes system with a damping term (we recall $\underline{u}_3^0 = 0$).

4.2.2. The “ill-prepared” case. We want here to present the result of [125] where $\Omega = \mathbb{T}^2 \times]0, h[$ and we consider “ill-prepared” initial data. Here, we have to study the oscillations in time and show that they do not affect the averaged flow. We can apply the same formal expansion as in the previous subsection taking into account the oscillations in time, namely

$$U = U^0 \left(\frac{t}{\varepsilon}, t, x, y, z, \frac{z}{l}, \frac{h-z}{l} \right) + \varepsilon U^1 + \dots, \quad (157)$$

$$U^0 = \underline{U}^0(\tau, t, x, y, z) + \tilde{U}^0(\tau, t, x, y, \theta) + \check{U}^0(\tau, t, x, y, \lambda). \quad (158)$$

We do not detail this expansion here and refer to [125]. We only point out that there are two extra difficulties here. Indeed, there is an oscillating boundary layer for each mode which has a vertical component. Moreover, we have to deal with the resonances between the different modes as in the works cited in the periodic case.

To write down the limit system, we introduce the spaces V_{sym}^s consisting of functions of H^s with some extra conditions on the boundary (see [125]). We also set $Lu = -P(e_3 \times u)$, where P is the projection onto divergence-free vector fields such that the third component vanishes on the boundary and $\mathcal{L}(\tau) = e^{\tau L}$. Let us denote w the solution in $L^\infty(0, T^*, V_{\text{sym}}^s)$ of the following system

$$\begin{cases} \partial_t w + \bar{Q}(w, w) - \Delta_{x,y} w + \gamma \bar{S}(w) = -\nabla p & \text{in } \Omega, \\ \operatorname{div} w = 0 & \text{in } \Omega, \\ w \cdot n = \pm w_3 = 0 & \text{on } \partial\Omega, \\ w(t=0) = w^0, \end{cases} \quad (159)$$

where $\bar{Q}(w, w)$, $\bar{S}(w)$ are respectively a bilinear and a linear operators of w , given by

$$\bar{Q}(w, w) = \sum_{\substack{l,m,k \\ k \in \mathcal{A}(l,m) \\ \lambda(l) + \lambda(m) = \lambda(k)}} b(t, l) b(t, m) \alpha_{lmk} N^k(X), \quad (160)$$

where the N^k are the eigenfunctions of L and $i\lambda(k)$ are the associated eigenvalues, α_{lmk} are constants which depends on (l, m, k) and $\mathcal{A}(l, m) = \{l + m, Sl + m, l + Sm, Sl + Sm\}$, ($Sl = (l_1, l_2, -l_3)$) is the set of possible resonances. The bilinear term $\bar{Q}$ is due to the fact that only resonant modes in the advective term $w \cdot \nabla w$ are present in the limit equation

$$\bar{S}(w) = \sum_k \frac{1}{h} (D(k) + iI(k)) b(t, k) N^k(X),$$

where

$$D(k) = \sqrt{2} \{ (1 - \lambda(k)^2)^{1/2} \}, \quad I(k) = \sqrt{2} \{ \lambda(k) (1 - \lambda(k)^2)^{1/2} \}.$$

In fact, $\bar{S}(w)$ is a damping term that depends on the frequencies $\lambda(k)$ since $D(k) \geq 0$. It is due to the presence of a boundary layer which creates a second flow of order ε responsible of this damping (called damping of Ekman).

THEOREM 4.1. *Let $s > 5/2$, and $w^0 \in V_{\text{sym}}^s(\Omega)^3$, $\nabla \cdot w^0 = 0$. We assume that u_0^n converges in $L^2(\Omega)$ to w^0 , $\eta = 1$ and ε, ν go to 0 such that $\sqrt{\nu/\varepsilon} \rightarrow \gamma$. Then any sequence of global weak solutions (à la Leray) u^n of (135)–(138) satisfying the energy inequality satisfies*

$$\begin{aligned} u^n - \mathcal{L}\left(\frac{t}{\varepsilon}\right) w &\rightarrow 0 & \text{in } L^\infty(0, T^*, L^2(\Omega)), \\ \nabla_{x,y} \left(u^n - \mathcal{L}\left(\frac{t}{\varepsilon}\right) w \right), \quad \sqrt{\nu} \partial_z u^n &\rightarrow 0 & \text{in } L^2(0, T^*, L^2(\Omega)), \end{aligned}$$

where w is the solution in $L^\infty(0, T^*, V_{\text{sym}}^s)$ of (159).

The above theorem gives a precise description of the oscillations in the sequence u^n . We can also show that the oscillations do not affect the averaged flow (also called the quasi-geotropic flow). We see then that $\bar{w}$ (the weak limit of u^n) satisfies a 2D Navier–Stokes equation with a damping term, namely

$$\begin{cases} \partial_t \bar{w} + \bar{w} \cdot \nabla \bar{w} - \eta \Delta_{x,y} \bar{w} + \gamma \frac{\sqrt{2}}{h} \bar{w} = -\nabla p & \text{in } \mathbb{T}^2, \\ \operatorname{div} \bar{w} = 0 & \text{in } \mathbb{T}^2, \\ \bar{w}(t=0) = \mathcal{S}(w^0) = \bar{w}^0, \end{cases} \quad (161)$$

where $\mathcal{S}$ is the projection onto the slow modes, namely that do not depend on z , $\bar{w}(t, x, y) = \mathcal{S}(w) = (1/h) \int_0^h w(t, x, y, z) dz$.

This can be proved by studying the operator Q and showing that if $k \in \mathcal{A}(l, m)$ with $k_3 = 0$ and $l_3 m_3 \neq 0$ then $\alpha_{lmk} + \alpha_{mlk} = 0$.

4.2.3. Nonflat bottom. In [125] we also deal with other boundary conditions, and construct Ekman layers near a non flat bottom

$$\Omega_\delta = \{(x, y, z), \text{ where } (x, y) \in \mathbb{T}^2, \text{ and } \delta f(x, y) < z < h\},$$

with the following boundary conditions

$$u(x, y, \delta f(x, y)) = 0. \quad (162)$$

We also treat the case of a free surface,

$$u_3^n(z=h) = 0, \quad \partial_z \begin{pmatrix} u_1^n \\ u_2^n \end{pmatrix} \Big|_{z=h} = \frac{1}{\beta} \sigma \left(\frac{t}{\varepsilon}, t, x, y \right), \quad (163)$$

where σ describes the wind (see [144]). Next, we have the following theorem.

THEOREM 4.2. *Let u^n be global weak solutions of (135)–(137), (162) and (163). If $\eta = 1$ and $(\varepsilon, \nu, \beta, \delta) \rightarrow (0, 0, 0, 0)$ then*

$$\begin{aligned} u^\nu - \mathcal{L} \left(\frac{t}{\varepsilon} \right) w &\rightarrow 0 && \text{in } L^\infty(0, T^*; L^2(\Omega)), \\ \nabla_{x,y} \left(u^\nu - \mathcal{L} \left(\frac{t}{\varepsilon} \right) w \right), &\quad \sqrt{\nu} \partial_z u^\nu \rightarrow 0 && \text{in } L^2(0, T^*; L^2(\Omega)), \end{aligned}$$

where w is the solution of the following system ($\sqrt{\nu/\varepsilon}$, ν/β , δ/ε stand for the limit of these quantities when n goes to infinity):

$$\begin{cases} \partial_t w + \bar{Q}(w, w) - \Delta_{x,y} w + \frac{1}{2} \sqrt{\frac{\nu}{\varepsilon}} \bar{S}(w) + \frac{\nu}{\beta} \bar{S}_1(\sigma) + \frac{\delta}{\varepsilon} \bar{S}_2(f, w) = -\nabla p, \\ \operatorname{div} w = 0 & \text{in } \Omega, \\ w \cdot n = \pm w_3 = 0 & \text{on } \partial\Omega, \\ w(t=0) = w^0, \end{cases} \quad (164)$$

where $\bar{S}_1(\sigma)$ and $\bar{S}_2(f, w)$ are source terms that are due respectively to the wind, and to the nonflat bottom.

The proofs of the above two theorems are based (as in the previous section) on energy estimates and use a more complicated corrector due to the presence of oscillations in time as well as the presence of different types of boundary layers. For more details about the proof, we refer to the original paper [125].

4.3. The case of other geometries

In the whole space case or in a domain $\Omega = \mathbb{R}^2 \times]0, h[$ the oscillations disperse to infinity as was the case for the acoustic waves in the compressible–incompressible limit. Let us state the following result for $\Omega = \mathbb{R}^2 \times]0, h[$ taken from [35]. We take η to be constant and $\nu = \varepsilon$.

THEOREM 4.3. *Let u_0 be a divergence free vector field is L^2 , $u_0 \cdot n = u_{03} = 0$ on $\partial\Omega$. Let u^ε be a family of weak solutions of (135)–(138) written in $\Omega = \mathbb{R}^2 \times]0, h[$. Let $\bar{w}$ be the global solution of the 2D Navier–Stokes system (161) in $\mathbb{R}^2$ with the initial data $S(w^0)$. Then we have*

$$\begin{aligned} & \|u^\varepsilon - (\bar{w}, 0)\|_{L^\infty(\mathbb{R}_+; L^2_{\text{loc}}(\mathbb{R}^2 \times]0, h[))} \\ & + \|\nabla(u^\varepsilon - (\bar{w}, 0))\|_{L^2(\mathbb{R}_+; L^2_{\text{loc}}(\mathbb{R}^2 \times]0, h[))} \rightarrow 0 \end{aligned} \quad (165)$$

when ε goes to zero.

The proof of this theorem uses the Ekman layer constructed in Section 4.2.1 and some Strichartz-type estimate for the oscillating part.

Let us also mention that the study of other geometries such as cylindrical domains were also studied [25].

4.4. Other related problems

We would like to end this section on rotating fluids by mentioning few related results. First, other physical systems present very similar properties to the rotating fluids. For instance there are several singular limits coming from magneto-hydrodynamic which have similar properties as the rotating fluids. We refer to [51] and [20].

An other important question concerns the stability of boundary layers. Indeed, in the previous subsection, we dealt with the case the horizontal viscosity was not going to zero. We can also study the case where η goes to zero. For the case without rotation we are lead to the inviscid limit which was studied in Section 2. It was proved that if ν , η and ν/η go to zero then we have convergence toward the Euler system. In other words the horizontal viscosity has a regularizing effect which is not shared by the vertical one. In the case with

rotation and when $\nu = \eta$, we can prove [123] (see also [125] for the ill-prepared case) that if

$$\|w\|_{L^\infty} \leq C \frac{\nu}{\varepsilon} \quad (166)$$

for some small enough constant C , then we have convergence toward the Euler system with damping, namely (161) with $\eta = 0$. This means that the rotation has a regularizing effect. Condition (166) is a stability condition. It was proved in [53] that the boundary layer can be unstable if (166) is not satisfied. More precisely, Desjardins and Grenier [53] prove the instability of the Ekman boundary layer under a more precise spectral condition. The stability condition (166) can also be refined to match the spectral condition. This was done by Rousset [148] for the case of Ekman boundary layers and [147] for the case of Ekman–Hartmann boundary layers.

5. Hydrodynamic limit of the Boltzmann equation

From a physical point of view, we expect that a gas can be described by a fluid equation when the mean free path (Knudsen number) goes to zero. During the last two decades this problem got a lot of interest and specially after DiPerna and Lions constructed their renormalized solutions [56]. In this section we present some of the most recent results concerning these (rigorous) derivations. We will present results for the three most classical equations of fluid mechanics in the incompressible regime, namely the incompressible Navier–Stokes equation, the Stokes equation and the Euler equation. We will also present some derivation of Fluid Mechanic boundary conditions starting from kinetic boundary conditions [132].

5.1. Scalings and formal asymptotics

In his sixth problem, Hilbert asked for a full mathematical justification of fluid mechanics equations starting from particle systems [88]. If we take the Boltzmann equation as a starting point, this problem can be stated as an asymptotic problem. Namely, starting from the Boltzmann equation, can we derive fluid mechanics equations and in which regime?

A program in this direction was initiated by Bardos, Golse and Levermore [12] who, using the renormalized solutions to the Boltzmann equation constructed by DiPerna and Lions, set an asymptotic regime where one can derive different fluid equations (and in particular incompressible models) depending on the chosen scaling.

5.1.1. The Boltzmann equation. The Boltzmann equation describes the evolution of the particle density of a rarefied gas. Indeed, the molecules of a gas can be modeled by hard spheres that move according to the laws of classical mechanics. However, due to the enormous number of molecules (about 2.7×10^{19} molecules in a cubic centimeter of gas at 1 atm and 0°C), it seems difficult to describe the state of the gas by giving the position and velocity of each individual particle. Hence, we must use some statistics and instead

of giving the position and velocity of each particle, we specify the density of particles $F(x, v)$ at each point x and velocity v . This means that we describe the gas by giving for each point x and velocity v the number of particles $F(x, v) dx dv$ in the volume $(x, x + dx) \times (v, v + dv)$.

Under some assumptions (rarefied gas, ...), it is possible to derive (at least formally) the Boltzmann equation from the classical Newton laws in an asymptotic regime where the number of particles goes to infinity (see [31,101,157] for some rigorous results about the derivation of the Boltzmann equation starting from the N particle system).

The Boltzmann equation reads

$$\partial_t F + v \cdot \nabla_x F = B(F, F), \quad (167)$$

where the collision kernel $B(F, F)$ is a quadratic form which acts only on the v variable. It describes the possible interaction between two different particles and is given by

$$B(F, F)(v) = \int_{\mathbb{R}^D} \int_{S^{D-1}} (F'_1 F' - F_1 F) b(v - v_1, \omega) dv_1 d\omega, \quad (168)$$

where we have used the following notation for all function ϕ

$$\phi' = \phi(v'), \quad \phi_1 = \phi(v_1), \quad \phi'_1 = \phi(v'_1) \quad (169)$$

and where the primed speeds are given by

$$v' = v + \omega[\omega \cdot (v_1 - v)], \quad v'_1 = v - \omega[\omega \cdot (v_1 - v)]. \quad (170)$$

Moreover, the Boltzmann cross-section $b(z, \omega)$, $z \in \mathbb{R}^D$, $\omega \in S^{D-1}$, depends on the molecular interactions (intermolecular potential). It is a nonnegative, locally integrable function (at least when grazing collisions are neglected). The Galilean invariance of the collisions implies that b depends only on $v - v_1$, ω and that

$$b(z, \omega) = |z| S(|z|, |\mu_c|), \quad \mu_c = \frac{\omega \cdot (v_1 - v)}{|v_1 - v|}, \quad (171)$$

where S is the specific differential cross-section. We also insist on the fact that the relations (170) are equivalent to the following conservations

$$v' + v'_1 = v + v_1 \quad (\text{conservation of the moment}), \quad (172)$$

$$|v'|^2 + |v'_1|^2 = |v|^2 + |v_1|^2 \quad (\text{conservation of the kinetic energy}). \quad (173)$$

We notice that the fact that two particles give two particles after the interaction translates the conservation of mass. For a more precise discussion about the Boltzmann equation, we refer to [30,31,168]. For some numerical works on the hydrodynamic limit, we refer to [156].

5.1.2. Compressible Euler. We start here by explaining how one can derive (at least formally) the Compressible Euler equation from the Boltzmann equation. A rigorous derivation can be found in Caflisch [28]. If F satisfies the Boltzmann equation, we deduce by integration in the v variable (at least formally) the following local conservations

$$\begin{cases} \partial_t \left(\int_{\mathbb{R}^D} F \, dv \right) + \nabla_x \cdot \left(\int_{\mathbb{R}^D} v F \, dv \right) = 0, \\ \partial_t \left(\int_{\mathbb{R}^D} v F \, dv \right) + \nabla_x \cdot \left(\int_{\mathbb{R}^D} v \otimes v F \, dv \right) = 0, \\ \partial_t \left(\int_{\mathbb{R}^D} |v|^2 F \, dv \right) + \nabla_x \cdot \left(\int_{\mathbb{R}^D} v |v|^2 F \, dv \right) = 0. \end{cases} \quad (174)$$

These three equations describe respectively the conservation of mass, momentum and energy. They present a great resemblance with the compressible Euler equation. However, the third moment $\int_{\mathbb{R}^D} v |v|^2 F \, dv$ is not a function of the others and depends in general on the whole distribution $F(v)$. In the asymptotic regimes we want to study, the distribution $F(v)$ will be very close to a Maxwellian due to the fact that the Knudsen number is going to 0. If we make the assumption that $F(v)$ is a Maxwellian for all t and x , then the third moment $\int_{\mathbb{R}^D} v |v|^2 F \, dv$ can be given as a function of $\rho = \int_{\mathbb{R}^D} F \, dv$, $\rho u = \int_{\mathbb{R}^D} v F \, dv$ and $\rho(|u|^2/2 + D\theta/2) = \int_{\mathbb{R}^D} \frac{1}{2} |v|^2 F \, dv$. Moreover, for all i and j , $\int_{\mathbb{R}^D} v_i v_j F \, dv$ can also be expressed as a function of ρ , u and θ .

We recall that a Maxwellian $M_{\rho,u,\theta}$ is completely defined by its density, bulk velocity and temperature,

$$M_{\rho,u,\theta} = \frac{\rho}{(2\pi\theta)^{D/2}} \exp\left(-\frac{1}{2\theta}|v-u|^2\right), \quad (175)$$

where ρ , u and θ depend only on t and x . If, we assume that for all t and x , F is a Maxwellian given by $F = M_{\rho(t,x),u(t,x),\theta(t,x)}$ then (174) reduces to

$$\begin{cases} \partial_t \rho + \nabla_x \cdot \rho u = 0, \\ \partial_t (\rho u) + \nabla_x \cdot (\rho u \otimes u) + \nabla_x (\rho \theta) = 0, \\ \partial_t \left(\frac{1}{2} \rho |u|^2 + \frac{D}{2} \rho \theta \right) + \nabla_x \cdot \left(\rho u \left(\frac{1}{2} |u|^2 + \frac{D+2}{2} \theta \right) \right) = 0, \end{cases} \quad (176)$$

which is the compressible Euler system for a monoatomic perfect gas. This derivation can become rigorous, if we take a sequence of solutions F_ε of

$$\partial_t F_\varepsilon + v \cdot \nabla_x F_\varepsilon = \frac{1}{\varepsilon} B(F_\varepsilon, F_\varepsilon), \quad (177)$$

where ε is the Knudsen number which goes to 0 (see [28]). Formally the presence of the term $\frac{1}{\varepsilon}$ in front of $\frac{1}{\varepsilon} B(F_\varepsilon, F_\varepsilon)$ implies (at the limit) that $B(F, F) = 0$ which means that F is a Maxwellian (see [30,31] or [168] for a proof of this fact).

5.1.3. Incompressible scalings. In the last subsection, we explained how we can derive the compressible Euler equation. It turns out that using different scalings, one can also derive incompressible models. We will explain what these scalings mean concerning the

Knudsen, Reynolds and Mach numbers. We consider the following global Maxwellian M which corresponds to $\rho = \theta = 1$ and $u = 0$.

$$M(v) = \frac{1}{(2\pi)^{D/2}} \exp\left(-\frac{1}{2}|v|^2\right). \quad (178)$$

Let $F_\varepsilon = MG_\varepsilon = M(1 + \varepsilon^m g_\varepsilon)$ be a solution of the following Boltzmann equation

$$\varepsilon^s \partial_t F_\varepsilon + v \cdot \nabla F_\varepsilon = \frac{1}{\varepsilon^q} B(F_\varepsilon, F_\varepsilon) \quad (179)$$

which is also equivalent to

$$\varepsilon^s \partial_t G_\varepsilon + v \cdot \nabla G_\varepsilon = \frac{1}{\varepsilon^q} Q(G_\varepsilon, G_\varepsilon), \quad (180)$$

where

$$Q(G, G)(v) = \int_{\mathbb{R}^D} \int_{S^{D-1}} (G'_1 G' - G_1 G) b(v - v_1, \omega) M_1 dv_1 d\omega. \quad (181)$$

With this scaling, we can define

$$Ma = \varepsilon^m, \quad Kn = \varepsilon^q, \quad Re = \varepsilon^{m-q}. \quad (182)$$

Here ε^s is a time scaling which is related to the Strouhal number. We recall that $St = L/(TU)$ and hence $St = \varepsilon^{s-m}$. This scaling in time allows us to choose the phenomenon we want to emphasize. By varying m, q and s , we can formally derive the following systems (see the references below for some rigorous mathematical results). A part from the first case where the compressible Euler system is satisfied by the moments of F , the fluid equations are recovered for the moments of the fluctuation g and we can show at least formally that $g = \rho + u \cdot v + \theta(|v|^2/2 - D/2)$ where (ρ, u, θ) satisfies one of the equations:

- (1) $q = 1, m = 0, s = 0$, compressible Euler system [28,100,167].
- (2) $q = 1, m > 0, s = 0$, acoustic waves [14]:

$$\begin{cases} \partial_t \rho + \nabla_x \cdot u = 0, \\ \partial_t u + \nabla_x(\rho + \theta) = 0, \\ \partial_t(\rho + \theta) + \frac{D+2}{D} \nabla_x \cdot u = 0. \end{cases} \quad (183)$$

We notice here that for these two first cases, we have $StMa = 1$ which is the condition to see some acoustic effects at the limit.

- (3) $q = 1, m = 1, s = 1$, incompressible Navier–Stokes–Fourier system [12,16,49,78, 114]:

$$\begin{cases} \partial_t u + u \cdot \nabla u - \nu \Delta u + \nabla p = 0, & \nabla_x \cdot u = 0, \\ \partial_t \theta + u \cdot \nabla \theta - \kappa \Delta \theta = 0, & \rho + \theta = 0. \end{cases}$$

(4) $q = 1, m > 1, s = 1$, Stokes–Fourier system [13,14,74,115,132]:

$$\begin{cases} \partial_t u - \nu \Delta u + \nabla p = 0, & \nabla_x \cdot u = 0, \\ \partial_t \theta - \kappa \Delta \theta = 0, & \rho + \theta = 0. \end{cases}$$

(5) $q > 1, m = 1, s = 1$, incompressible Euler–Fourier system [115,149]:

$$\begin{cases} \partial_t u + u \cdot \nabla u + \nabla p = 0, & \nabla_x \cdot u = 0, \\ \partial_t \theta + u \cdot \nabla \theta = 0, & \rho + \theta = 0. \end{cases}$$

Note that the compressible Navier–Stokes system (with a viscosity of order 1) cannot be derived in this manner because of the following physical relation

$$Re = C \frac{Ma}{Kn}. \quad (184)$$

However, the compressible Navier–Stokes system with a viscosity of order ε can be considered as a better approximation than the compressible Euler system in the case $q = 1, m = 0, s = 0$.

5.1.4. Formal development. Here, we want to explain (at least formally) how we can derive the incompressible Navier–Stokes system for the bulk velocity and the Fourier equation for the temperature starting from the Boltzmann system with the scalings $q = 1, m = 1, s = 1$. A simple adaptation of the argument also yields a formal derivation of the Stokes–Fourier system (which is the linearization of the Navier–Stokes–Fourier system) as well as the Euler. Rewriting the equation satisfied by g_ε , we get

$$\partial_t g_\varepsilon + \frac{1}{\varepsilon} v \cdot \nabla_x g_\varepsilon = -\frac{1}{\varepsilon^2} L g_\varepsilon + \frac{1}{\varepsilon} Q(g_\varepsilon, g_\varepsilon), \quad (185)$$

where L is the linearized collision operator given by

$$Lg = \int_{\mathbb{R}^D} \int_{S^{D-1}} (g + g_1 - g'_1 - g') b(v - v_1, \omega) M_1 dv_1 d\omega. \quad (186)$$

We assume that g_ε can be decomposed as follows $g_\varepsilon = g + \varepsilon h + \varepsilon^2 k + O(\varepsilon^3)$ and we make the following formal development

$$\frac{1}{\varepsilon^2} : Lg = 0. \quad (187)$$

A simple study of the operator L shows that it is formally self-adjoint, nonnegative for the following scalar product $\langle f, g \rangle = \langle fg \rangle$ where we use the following notation $\langle g \rangle = \int_{\mathbb{R}^D} g M dv$ and $\text{Ker}(L) = \{g, g = \alpha + \beta \cdot v + \gamma |v|^2, \text{ where } (\alpha, \beta, \gamma) \in \mathbb{R} \times \mathbb{R}^D \times \mathbb{R}\}$. Hence, we deduce that $g = \rho + u \cdot v + \theta(|v|^2/2 - D/2)$.

$$\frac{1}{\varepsilon} : v \cdot \nabla g = -Lh + Q(g, g). \quad (188)$$

Integrating over v , we infer that $u = \langle vg \rangle$ is divergence-free ($\operatorname{div} u = 0$). Moreover, multiplying by v and taking the integral over v , we infer that $\nabla(\rho + \theta) = 0$ which is the Boussinesq relation. Besides, at order 1, we have

$$\frac{1}{\varepsilon_0} : \partial_t g + v \cdot \nabla_x h = -Lk + 2Q(g, h), \quad (189)$$

from which we deduce that

$$\frac{1}{\varepsilon_0} : \partial_t \langle vg \rangle + \nabla_x \cdot \langle v \otimes vh \rangle = 0, \quad (190)$$

$$\frac{1}{\varepsilon_0} : \partial_t \left\langle \left(\frac{|v|^2}{D+2} - 1 \right) g \right\rangle + \nabla_x \cdot \left\langle v \left(\frac{|v|^2}{D+2} - 1 \right) h \right\rangle = 0. \quad (191)$$

To get a closed equation for g , we have to inverse the operator L . We define the matrix $\phi(v)$ and the vector $\psi(v)$ as the unique solutions of

$$L\phi(v) = v \otimes v - \frac{1}{D} |v|^2 I, \quad L\psi(v) = \left(\frac{|v|^2}{D+2} - 1 \right) v \quad (192)$$

which are orthogonal to $\operatorname{Ker}(L)$ for the scalar product $\langle \cdot, \cdot \rangle$. We also define the viscosity ν and the heat conductivity κ by

$$\nu = \frac{1}{(D-1)(D+2)} \langle \phi : L\phi \rangle, \quad (193)$$

$$\kappa = \frac{2}{D(D+2)} \langle \psi \cdot L\psi \rangle. \quad (194)$$

We notice that ν and κ only depend on b . Using that L is formally self-adjoint, we deduce that

$$\partial_t \langle gv_i \rangle + \nabla_x \cdot \langle \phi_{ij} (Q(g, g) - v \cdot \nabla g) \rangle + \nabla \left\langle \frac{|v|^2}{N} h \right\rangle = 0, \quad (195)$$

$$\partial_t \left\langle g \left(\frac{|v|^2}{D+2} - 1 \right) \right\rangle + \nabla_x \cdot \langle \psi (Q(g, g) - v \cdot \nabla g) \rangle = 0. \quad (196)$$

A simple (but long) computation gives the Navier–Stokes equation and the Fourier equation, namely

$$\partial_t u + u \cdot \nabla u - \nu \Delta u + \nabla p = 0, \quad (197)$$

$$\partial_t \theta + u \cdot \nabla \theta - \kappa \Delta \theta = 0, \quad (198)$$

where $u = \langle gv \rangle$, $\theta = -\rho = \langle (|v|^2/(D+2) - 1)g \rangle$ and the pressure p is the sum of different contributions.

5.1.5. Mathematical difficulties. Here, we want to explain the major mathematical difficulties encountered in trying to give a rigorous justification of any of the above asymptotic problems starting from renormalized solutions.

- D1. The local conservation of momentum is not known to hold for the renormalized solutions of the Boltzmann equation. Indeed, the solutions constructed by DiPerna and Lions [56] only hold in the renormalized sense which means that

$$\partial_t \beta(F) + v \cdot \nabla \beta(F) = Q(F, F) \beta'(F), \quad (199)$$

$$\beta(F)(t=0) = \beta(F^0), \quad (200)$$

where β is given, for instance, by $\beta(f) = \log(1 + f)$.

- D2. The lack of a priori estimates. Indeed, all we can deduce from the entropy inequality and the conservation of energy is that g_ε is bounded in $L \log L$ and that $g_\varepsilon |v|^2$ is bounded in L^1 . However, we need a bound in L^2 to define all the product involved in the formal development. In [74], the authors used the entropy dissipation estimate to deduce some information on the structure of the fluctuation g_ε and get some new a priori estimates by using some Caflisch–Grad estimates.

To pass to the limit in the different products (and specially in the case we want to recover the Navier–Stokes–Fourier system or the Euler system), one has also to prove that g_ε is compact in space and time, namely that $g_\varepsilon \in K$ where K is a compact subset of some $L^p(0, T; L^1(\Omega))$. We split this in three difficulties.

- D3. The compactness in space of g_ε . This was achieved in the stationary case by Bardos, Golse and Levermore [12,15], using averaging lemma [75,76] and proving that g_ε is in some compact subset of $L^1(\Omega)$. However, a newer version of the averaging lemma [78] was needed in [79] to prove some equiintegrability and hence the absence of concentration.
- D4. The compactness in time for g_ε . It turns out that in general g_ε is not compact in time. Indeed, g_ε presents some oscillations in time which can be analyzed and described precisely. Using this description and some compensation (due to a remarkable identity satisfied by the solutions to the wave equation), it is possible to pass to the limit in the whole equation. This was done by Lions and the author [114] using some ideas coming from the compressible–incompressible limit [111,113].
- D5. An other difficulty is that in [12], very restrictive conditions on the Boltzmann kernel were imposed. These conditions were slightly relaxed in [74] to treat some general hard potentials in the Stokes–Fourier scaling and in [79] to treat Maxwellian potential. The case of general potentials including soft potentials was treated in [105].

5.2. The convergence toward the incompressible Navier–Stokes–Fourier system

The first paper dealing with the rigorous justification of the formal development Section 5.1.4 goes back to the work of Bardos, Golse and Levermore [12] where the stationary case was handled under different assumptions and restrictions (see also De Masi, Esposito

and Lebowitz [49] for a similar result in a different setting). There are however some aspects of the analysis performed in [12] that can be improved. First, the heat equation was not treated because the heat flux terms could not be controlled. Second, local momentum conservation was assumed because DiPerna–Lions solutions are not known to satisfy the local conservation law of momentum (or energy) that one would formally expect. Third, the discrete-time case was treated in order to avoid having to control the time regularity of the acoustic modes. Fourth, unnatural technical assumptions were made on the Boltzmann kernel. Finally, a mild compactness assumption was required to pass to the limit in certain nonlinear terms.

During the last few years, there appeared several results trying to improve the result of [12] and give a rigorous justification of the derivation. In [114] and under two assumptions (the conservation of the momentum and a compactness assumption), it was possible to treat the time dependent case and derive the incompressible Navier–Stokes equation. In [74] Golse and Levermore gave a rigorous derivation of Stokes–Fourier system (the linearization of the Navier–Stokes–Fourier system) without any assumption. In [79] Golse and Saint-Raymond gave the first derivation of the Navier–Stokes–Fourier system without any compactness or momentum assumption. However, their result only applies to a small class of collision kernels. In a recent work in collaboration with Levermore [105], we give a derivation of the Navier–Stokes–Fourier system for a very general class of Boltzmann kernels which includes in particular soft potentials.

In what follows, we assume that Ω is the whole space or the torus to avoid dealing with the boundary. First, let us specify the conditions we impose on the initial data. It is supposed that G_ε^0 satisfies (we recall that $F_\varepsilon^0 = MG_\varepsilon^0$)

$$H(G_\varepsilon^0) = \int_{\Omega} \int_{\mathbb{R}^D} (G_\varepsilon^0 \log G_\varepsilon^0 - G_\varepsilon^0 + 1) M \, dx \, dv \leq C\varepsilon^2. \quad (201)$$

This shows that we can extract a subsequence of the sequence g_ε^0 (defined by $G_\varepsilon^0 = 1 + \varepsilon g_\varepsilon^0$) which converges weakly in L^1 toward g^0 such that $g^0 \in L^2$. We also notice that (201) is equivalent to the fact that $\int_{\Omega} \langle h(\varepsilon g_\varepsilon^0) \rangle \, dx \leq C\varepsilon^2$, where $h(z) = (1+z) \log(1+z) - z$ which is almost an L^2 estimate for g_ε^0 . This shows at least that $g^0 \in L^2$. Then, we consider a sequence G_ε of renormalized solutions of the Boltzmann equation (180) with $s = q = 1$, satisfying the entropy inequality and we want to prove that g_ε converges to some $g = u \cdot v + \theta(|v|^2/2 - (D+2)/2)$.

Before stating the new result of Golse and Saint-Raymond [77], we want to explain the kind of assumptions that were made in previous works. The convergence result proved in [114] (which only deals with the u component) requires the following two hypotheses (A1) and (A2) on the sequence G_ε which allow to circumvent the difficulties D1 and D2.

(A1) The solution G_ε satisfies the projection on divergence-free vector fields of the local momentum conservation law

$$\partial_t P \langle v G_\varepsilon \rangle + \frac{1}{\varepsilon} P \nabla_x \cdot \langle v \otimes v G_\varepsilon \rangle = 0. \quad (202)$$

(A2) The family $(1 + |v|^2)g_\varepsilon^2/N_\varepsilon$ is relatively compact for the weak topology of $L^1(dt M \, dv \, dx)$ which we denote $w - L^1(dt M \, dv \, dx)$, where $N_\varepsilon = 1 + \frac{\varepsilon}{3}g_\varepsilon$.

In the sequel, we denote the weak topology of $L^1(dt M dv dx)$ by $w - L^1(dt M dv dx)$. The assumption (A2) enforces the $L \log L$ estimate we have on g_ε , namely $\int_\Omega \langle h(\varepsilon g_\varepsilon) \rangle dx \leq C\varepsilon^2$ to prevent some type of concentration.

Now, we state the result of Golse and Saint-Raymond [79] where no assumptions on the solutions is made. This result was extended by Levermore and the author [105] to treat the case of a larger class of Boltzmann kernels which includes all the classical kernels in particular soft potentials.

Under some assumptions on the Boltzmann kernel (see [79,105]), we have the theorem.

THEOREM 5.1. *Let G_ε be a sequence of renormalized solutions of the Boltzmann equations (180) with initial condition G_ε^0 and satisfying the entropy inequality. Then, the family $(1 + |v|^2)g_\varepsilon$ is relatively compact in $w - L^1(dt M dv dx)$. If g is a weak limit of a subsequence (still denoted g_ε) then $Lg = 0$ and $g = \rho + u \cdot v + \theta(|v|^2/2 - D/2)$ satisfies the limiting dissipation inequality*

$$\begin{aligned} & \frac{1}{2} \int_\Omega |\rho(t)|^2 + |u(t)|^2 + \frac{D}{2} |\theta(t)|^2 dx \\ & + \int_0^t \int_\Omega \frac{1}{2} v |\nabla_x u + {}^t \nabla_x u|^2 + \kappa |\nabla \theta|^2 \\ & \leq \liminf_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \int_\Omega \langle h(\varepsilon g_\varepsilon) \rangle dx = C^0. \end{aligned} \quad (203)$$

Moreover, $\theta + \rho = 0$ and $(u, \theta) = (\langle v g \rangle, \langle (|v|^2/(D+2) - 1)g \rangle)$ is a weak solution of the Navier–Stokes–Fourier system (NSF):

$$\begin{cases} \partial_t u + u \cdot \nabla u - v \Delta u + \nabla p = 0, & \nabla \cdot u = 0, \\ \partial_t \theta + u \cdot \nabla \theta - \kappa \Delta \theta = 0, \\ u(t=0, x) = u^0(x), & \theta(t=0, x) = \theta^0(x), \end{cases} \quad (\text{NSF})$$

with the initial condition $u^0 = P\langle v g^0 \rangle$ and $\theta^0 = \langle (|v|^2/(D+2) - 1)g^0 \rangle$ and where the viscosity v and heat conductivity κ are given by (193) and (194).

IDEA OF THE PROOF. Now, we give an idea of the proof of Theorem 5.1 (see [79] and [105] for a complete proof). We start by recalling a few a priori estimates taken from [12].

PROPOSITION 5.2. *We have:*

(i) *The sequence $(1 + |v|^2)g_\varepsilon$ is bounded in $L^\infty(dt; L^1(M dv dx))$ and relatively compact in $w - L^1(dt M dv dx)$. Moreover, if g is the weak limit of any converging subsequence of g_ε , then $g \in L^\infty(dt; L^2(M dv dx))$ and for almost every $t \in [0, \infty)$, we have*

$$\frac{1}{2} \int_\Omega \langle g^2(t) \rangle dx \leq \liminf_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \int_\Omega \langle h(\varepsilon g_\varepsilon(t)) \rangle dx \leq C^0. \quad (204)$$

(ii) Denoting $q_\varepsilon = \frac{1}{\varepsilon^2}(G'_{\varepsilon 1}G'_\varepsilon - G_{\varepsilon 1}G_\varepsilon)$, we have that the sequence $(1 + |v|^2)q_\varepsilon/N_\varepsilon$ is relatively compact in $w - L^1(dt d\mu dx)$, where $d\mu = b(v - v_1, \omega) d\omega M_1 dv_1 M dv$. Besides, if q is the weak limit of any converging subsequence of $q_\varepsilon/N_\varepsilon$ then $q \in L^2(dt; L^2(d\mu dx))$ and q inherits the same symmetries as q_ε , namely $q(v, v_1, \omega) = q(v_1, v, \omega) = -q(v', v'_1, \omega)$.

(iii) In addition, for almost all (t, x) , $Lg = 0$, which means that g is of the form

$$g(t, x, v) = \rho(t, x) + u(t, x) \cdot v + \theta(t, x) \left(\frac{1}{2}|v|^2 - \frac{D}{2} \right), \quad (205)$$

where $\rho, u, \theta \in L^\infty(dt; L^2(dx))$.

(iv) Finally, from the renormalized equation, we deduce that

$$v \cdot \nabla_x g = \int \int q b(v_1 - v, \omega) d\omega M_1 dv_1 \quad (206)$$

which yields the incompressibility and Boussinesq relations, namely

$$\nabla_x \cdot u = 0, \quad \nabla_x(\rho + \theta) = 0. \quad (207)$$

The rest of the proof is based on a new averaging lemma [78] as well as a better use of the entropy dissipation to get some estimate on the non hydrodynamic part of g_ε . The final passage to the limit uses the same local method of Section 3.3.1 to deal with the acoustic waves. $\square$

REMARK 5.3. Let us also mention a new work of Guo [86] where he proves that the next order terms in the formal development also hold for the case of regular solutions to the Boltzmann equation.

5.3. The convergence toward the Stokes system

The convergence toward the Stokes system is easier than the Navier–Stokes case for two reasons. Indeed, we do not have to pass to the limit in the nonlinear terms. Besides, the control we get from the entropy dissipation is better. In this section, we want to present the result of [115] where a new notion of renormalized solution was used. In [74], the whole Stokes–Fourier system was also recovered by using a different method.

5.3.1. Defect measures. In [115] the difficulty D1 was overcome by showing that the conservation of momentum can be recovered in the limit by a very simple argument. Indeed by looking at the construction of the renormalized solutions of DiPerna and Lions [56], one sees that one can write a kind of conservation of moment (with a defect measure) which also intervenes in the energy inequality. Indeed, the solutions F_ε built by DiPerna and Lions satisfy in addition

$$\partial_t \int_{\mathbb{R}^D} v F_\varepsilon dv + \frac{1}{\varepsilon} \operatorname{div} \int_{\mathbb{R}^D} (v \otimes v) F_\varepsilon dv + \frac{1}{\varepsilon} \operatorname{div}(M_\varepsilon) = 0. \quad (208)$$

Besides, the following energy equality holds

$$\frac{1}{2} \int_{\Omega} \int_{\mathbb{R}^D} |v|^2 F_{\varepsilon}(t, x, v) dx dv + \frac{1}{2} \int_{\Omega} \operatorname{tr}(M_{\varepsilon}) dx = \frac{1}{2} \int_{\Omega} \int_{\mathbb{R}^D} |v|^2 F_{\varepsilon}^0(x, v) dx dv \quad (209)$$

which can be rewritten (with $\varepsilon^m m_{\varepsilon} = M_{\varepsilon}$),

$$\partial_t \langle g_{\varepsilon} v \rangle + \nabla_x \cdot \langle g_{\varepsilon} v \otimes v \rangle + \frac{1}{\varepsilon} \nabla \cdot m_{\varepsilon} = 0, \quad (210)$$

$$\int_{\Omega} \langle |v|^2 g_{\varepsilon} \rangle dx + \int_{\Omega} \operatorname{tr}(m_{\varepsilon}) dx = 0. \quad (211)$$

5.3.2. Entropy inequality. One can write the entropy inequality for G_{ε} (as in the case of the limit toward the Navier–Stokes system) or write it for F_{ε} as well. It turns out that the second choice gives a better estimate for the defect measure. Indeed starting from the entropy inequality for F_{ε} , we can deduce

$$\begin{aligned} & \int_{\Omega} \int_{\mathbb{R}^D} h(\varepsilon^m g_{\varepsilon}) dx M dv(t) - \int_{\Omega} \int_{\mathbb{R}^D} \varepsilon^m \frac{|v|^2}{2} g_{\varepsilon} dx M dv(t) \\ & + \frac{1}{4\varepsilon^2} \int_0^t ds \int_{\Omega} dx \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} M dv M_1 dv_1 \int_{S^{D-1}} d\omega b(v - v_1, \omega) \\ & \times (G'_{\varepsilon 1} G'_{\varepsilon} - G_{\varepsilon 1} G_{\varepsilon}) \log \left(\frac{G'_{\varepsilon 1} G'_{\varepsilon}}{G_{\varepsilon 1} G_{\varepsilon}} \right) \\ & \leq \int_{\Omega} \int_{\mathbb{R}^D} h(\varepsilon^m g_{\varepsilon}^0) dx M dv. \end{aligned} \quad (212)$$

Let us now state the result. We take initial data satisfying

$$\int_{\mathbb{T}^D} \int_{\mathbb{R}^D} F_{\varepsilon}^0 dx dv = 1, \quad \int_{\mathbb{T}^D} \int_{\mathbb{R}^D} v F_{\varepsilon}^0 dx dv = 0, \quad \int_{\mathbb{T}^D} \int_{\mathbb{R}^D} |v|^2 F_{\varepsilon}^0 dx dv = D \quad (213)$$

and

$$\int_{\Omega} \int_{\mathbb{R}^D} F_{\varepsilon}^0 \log F_{\varepsilon}^0 dx dv \leq -\frac{D}{2} + C\varepsilon^{2m}. \quad (214)$$

We also assume that b satisfies (A0).

THEOREM 5.4. *If F_{ε} is a sequence of renormalized solutions of the Boltzmann equations (179), $s = q = 1$ and $m > 1$, with initial condition F_{ε}^0 and satisfies the entropy inequality*

as well as the refined momentum equation, then the family $(1 + |v|^2)g_\varepsilon$ is relatively compact in $w - L^1(dt M dv dx)$. And, if g is a weak limit of a subsequence (still denoted g_ε) then $Lg = 0$ and $g = \rho + u \cdot v + \theta(|v|^2/2 - N/2)$ satisfies the limiting dissipation inequality

$$\begin{aligned} & \frac{1}{2} \int_{\Omega} |\rho(t)|^2 + |u(t)|^2 + \frac{D}{2} |\theta|^2 dx + \int_0^t \int_{\Omega} \frac{1}{2} v |\nabla_x u + {}^t\nabla_x u|^2 \\ & \leq \liminf_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^{2m}} \int_{\Omega} \langle h(\varepsilon^m g_\varepsilon) \rangle dx = C^0. \end{aligned} \quad (215)$$

Moreover, $u = \langle vg \rangle$ is the solution of the Stokes system (S) with the initial condition $u^0 = P \langle vg^0 \rangle$ and where the viscosity ν is given by (193). Besides, we have the following strong Boussinesq relationship

$$\rho + \theta = 0. \quad (216)$$

We only explain here briefly how we can recover the conservation of momentum at the limit. Indeed, starting from the entropy inequality, one deduces that

$$\int_{\Omega} \langle h(\varepsilon^m g_\varepsilon) \rangle dx + \varepsilon^m \operatorname{tr}(m_\varepsilon) + D(G_\varepsilon) \leq C \varepsilon^{2m} \quad (217)$$

and since $m > 1$, we deduce

$$\frac{1}{\varepsilon} \operatorname{tr}(m_\varepsilon) \quad \text{and} \quad \frac{1}{\varepsilon} m_\varepsilon \rightarrow 0 \quad (218)$$

in $L^\infty(0, T; L^1(\Omega))$ since m_ε is a positive matrix. This yields the local conservation of momentum in (210) at the limit.

5.4. The case of a bounded domain

In this subsection we want to present the derivation of fluid mechanics boundary conditions starting from kinetic boundary condition. For simplicity, we will present the result in the Stokes scaling though the proof works as well for the Navier–Stokes scaling using the result of the previous sections. We also refer to [27] for a derivation of the Navier condition for the primitive equations.

Let Ω be a smooth bounded domain of $\mathbb{R}^D$ and $\mathcal{O} = \Omega \times \mathbb{R}^D$ the space-velocity domain. Let $n(x)$ be the outward unit normal vector at $x \in \partial\Omega$. We denote by $d\sigma_x$ the Lebesgue measure on the boundary $\partial\Omega$ and we define the outgoing/incoming sets Σ_+ and Σ_- by

$$\Sigma_{\pm} = \{(x, v) \in \Sigma, \pm n(x) \cdot v > 0\}, \quad \text{where } \Sigma = \partial\Omega \times \mathbb{R}^D.$$

We consider the Boltzmann equation in $\mathbb{R}_+ \times \mathcal{O}$ with a scaling where $q = s = 1$ and $m > 1$.

5.4.1. The Maxwell boundary condition. The boundary condition we will consider express the balance between the incoming and outgoing part of the trace of F , namely $\gamma_{\pm} F = \mathbb{1}_{\Sigma_{\pm}} \gamma F$. We will use the following Maxwell reflection condition

$$\gamma_- F = (1 - \alpha)L(\gamma_+ F) + \alpha K(\gamma_+ F) \quad \text{on } \Sigma_-, \quad (219)$$

where α is a constant also called accommodation coefficient. The local reflection operator L is given by

$$L\phi(x, v) = \phi(x, R_x v), \quad (220)$$

where $R_x v = v - 2(n(x) \cdot v)n(x)$ is the velocity before the collision with the wall. The diffuse reflection operator K is given by

$$K\phi(x, v) = \sqrt{2\pi}\tilde{\phi}(x)M(v), \quad (221)$$

where $\tilde{\phi}$ is the outgoing mass flux

$$\tilde{\phi}(x) = \int_{v \cdot n(x) > 0} \phi(x, v) n(x) \cdot v \, dv. \quad (222)$$

We notice that

$$\int_{v \cdot n(x) > 0} n(x) \cdot v \sqrt{2\pi} M(v) \, dv = \int_{v \cdot n(x) < 0} |n(x) \cdot v| \sqrt{2\pi} M(v) \, dv = 1,$$

which expresses the conservation of mass at the boundary. Here, we are taking the temperature of the wall to be constant and equal to 1. For the existence of renormalized solutions to the Boltzmann equation in a bounded domain we refer to [136].

5.4.2. A priori estimate. Let $\mathcal{E}(\gamma_+ G_{\varepsilon})$, the so-called Darrozès–Guiraud information [85], be given by

$$\mathcal{E}(\gamma_+ G_{\varepsilon}) = \int_{\partial\Omega} \left(\langle h(\delta_{\varepsilon} \gamma_+ g_{\varepsilon}) \rangle_{\partial\Omega} - h(\langle \delta_{\varepsilon} \gamma_+ g_{\varepsilon} \rangle_{\partial\Omega}) \right) d\sigma_x. \quad (223)$$

In the case of a bounded domain, the entropy inequality reads

$$H(G_{\varepsilon}(t)) + \int_0^t \left(\frac{1}{\varepsilon^2} E(G_{\varepsilon}(s)) + \frac{\alpha_{\varepsilon}}{\sqrt{2\pi\varepsilon}} \mathcal{E}_{\varepsilon}(\gamma_+ G_{\varepsilon}(s)) \right) ds \leq H(G_{\varepsilon}^{\text{in}}), \quad (224)$$

where $H(G)$ is the relative entropy functional

$$H(G) = \int_{\Omega} \langle (G \log(G) - G + 1) \rangle dx, \quad (225)$$

and $E(G)$ is the entropy dissipation rate functional

$$E(G) = \int_{\Omega} \left\langle \frac{1}{4} \log \left(\frac{G'_1 G'}{G_1 G} \right) (G'_1 G' - G_1 G) \right\rangle dx. \quad (226)$$

Notice the presence of the extra positive term due to the boundary. It is easy to see that due to Jensen inequality the extra term $\mathcal{E}_{\varepsilon}(\gamma_+ G_{\varepsilon}(s)) \geq 0$. This also gives a bound on $\gamma_+ G_{\varepsilon}$ which is useful.

Now, we present two results taken from [132] which hold for a wide range of collision kernels

THEOREM 5.5 (Navier boundary condition). *Let $F_{\varepsilon}^{\text{in}} = G_{\varepsilon}^{\text{in}} M$ be a family of initial data satisfying*

$$\frac{1}{\delta_{\varepsilon}^2} H(G_{\varepsilon}^{\text{in}}) + \int \int_O |v|^2 F_{\varepsilon}^{\text{in}} dx dv \leq C^{\text{in}} \quad (227)$$

for some $C^{\text{in}} < \infty$ and

$$\begin{aligned} \frac{1}{\delta_{\varepsilon}} \Pi \langle v G_{\varepsilon}^{\text{in}} \rangle &\rightarrow u && \text{in } \mathcal{D}'(\Omega; \mathbb{R}^D), \\ \frac{1}{\delta_{\varepsilon}} \left\langle \left(\frac{1}{D+2} |v|^2 - 1 \right) G_{\varepsilon}^{\text{in}} \right\rangle &\rightarrow \theta && \text{in } \mathcal{D}'(\Omega; \mathbb{R}^D), \end{aligned} \quad (228)$$

for some $(u^{\text{in}}, \theta^{\text{in}}) \in L^2(dx; \mathbb{R}^D \times \mathbb{R})$. Denote by G_{ε} any corresponding family of renormalized solutions of the Boltzmann equation satisfying the entropy inequality (224), where the accommodation coefficient satisfies

$$\frac{\alpha_{\varepsilon}}{\sqrt{2\pi\varepsilon}} \rightarrow \lambda \quad \text{when } \varepsilon \rightarrow 0. \quad (229)$$

Then, as $\varepsilon \rightarrow 0$, the family of fluctuations satisfies

$$\begin{aligned} g_{\varepsilon} &\rightarrow v \cdot u + \left(\frac{1}{2} |v|^2 - \frac{D+2}{2} \right) \theta \\ &\quad \text{in } w - L^1_{\text{loc}}(dt; w - L^1((1 + |v|^2) M dv dx)), \\ \Pi(v g_{\varepsilon}) &\rightarrow u && \text{in } C([0, \infty); \mathcal{D}'(\Omega; \mathbb{R}^D)), \\ \left\langle \left(\frac{1}{D+2} |v|^2 - 1 \right) g_{\varepsilon} \right\rangle &\rightarrow \theta && \text{in } C([0, \infty); \mathcal{D}'(\Omega; \mathbb{R}^D)), \end{aligned} \quad (230)$$

where Π is the orthogonal projection from $L^2(dx; \mathbb{R}^D)$ onto divergence-free vector fields with zero normal velocity, namely the set

$$H = \{u \in L^2(\Omega), \nabla_x \cdot u = 0, u \cdot n = 0 \text{ on } \partial\Omega\}.$$

Furthermore, $(u, \theta) \in C([0, \infty); H \times L^2(\Omega)) \cap L^2(dt; H^1(\Omega) \times H^1(\Omega))$ and it satisfies the Stokes–Fourier system with Navier boundary condition

$$\begin{cases} \partial_t u + \nabla_x p - \nu \Delta_x u = 0, & \operatorname{div} u = 0 & \text{on } \mathbb{R}^+ \times \Omega, \\ (2\nu d(u) \cdot n + \lambda u) \wedge n = 0, & u \cdot n = 0 & \text{on } \mathbb{R}^+ \times \partial\Omega, \\ \partial_t \theta - \kappa \Delta_x \theta = 0 & & \text{on } \mathbb{R}^+ \times \Omega, \\ \kappa \partial_n \theta + \lambda \frac{D+1}{D+2} \theta = 0 & & \text{on } \mathbb{R}^+ \times \partial\Omega, \end{cases} \quad (231)$$

$$u(0, x) = u^{\text{in}}(x), \quad \theta(0, x) = \theta^{\text{in}}(x) \quad \text{on } \Omega,$$

where $d(u)$ denotes the symmetric part of the stress tensor $d(u) = \frac{1}{2}(\nabla u + {}^t \nabla u)$.

The second result treats the case of Dirichlet boundary conditions. We will make the same assumptions as in the previous theorem but instead of assuming that $\alpha_\varepsilon/(\varepsilon\sqrt{2\pi}) \rightarrow \lambda$, we assume that $\alpha_\varepsilon/\varepsilon \rightarrow +\infty$.

THEOREM 5.6 (Dirichlet boundary condition). *We make the same assumptions as in Theorem 5.5, except that we replace condition (229) by*

$$\frac{\alpha_\varepsilon}{\varepsilon} \rightarrow \infty \quad \text{when } \varepsilon \rightarrow 0. \quad (232)$$

Then, as $\varepsilon \rightarrow 0$, we have the same convergences (230) as in Theorem 5.5 with $(u, \theta) \in C([0, \infty); H \times L^2(\Omega)) \cap L^2(dt; V \times H_0^1(\Omega))$, where

$$V = \{u \in H^1(\Omega), \nabla_x \cdot u = 0, u = 0 \text{ on } \partial\Omega\}.$$

Furthermore, (u, θ) satisfies the Stokes–Fourier system with Dirichlet boundary condition

$$\begin{cases} \partial_t u + \nabla_x p - \nu \Delta_x u = 0, & \operatorname{div} u = 0 & \text{on } \mathbb{R}^+ \times \Omega, \\ \partial_t \theta - \kappa \Delta_x \theta = 0 & & \text{on } \mathbb{R}^+ \times \Omega, \\ u = 0, & \theta = 0 & \text{on } \mathbb{R}^+ \times \partial\Omega, \\ u(0, x) = u^{\text{in}}(x), & \theta(0, x) = \theta^{\text{in}}(x) & \text{on } \Omega. \end{cases} \quad (233)$$

IDEA OF THE PROOF. The interior convergence can be deduced easily from the work of Golse and Levermore [74]. We just want to explain the convergence at the boundary. We prove two types of control on the trace γg_ε of g_ε on the boundary. The first control comes from the inside, it uses the interior estimates to deduce an estimate on the trace. $\square$

LEMMA 5.7. *We have for all $p > 0$,*

$$\gamma \hat{g}_\varepsilon \rightarrow \gamma g \quad \text{in } w - L_{\text{loc}}^1(dt; w - L^1(M(1 + |v|^p)|v \cdot n(x)| dv d\sigma_x)), \quad (234)$$

$$\varepsilon^m \gamma g_\varepsilon \rightarrow 0 \quad \text{a.e. on } \mathbb{R}^+ \times \partial\Omega \times \mathbb{R}^d. \quad (235)$$

The second control comes from the boundary term appearing in the entropy dissipation. It does not give an estimate on g_ε but rather on g_ε minus its average in v . We get the lemma.

LEMMA 5.8. *Define $\gamma_\varepsilon = \gamma_+ g_\varepsilon - \mathbb{1}_{\Sigma_+} \langle \gamma_+ g_\varepsilon \rangle_{\partial\Omega}$ and*

$$\gamma_\varepsilon^{(1)} = \gamma_\varepsilon \mathbb{1}_{\gamma_+ G_\varepsilon \leq 2 \langle \gamma_+ G_\varepsilon \rangle_{\partial\Omega} \leq 4 \gamma_+ G_\varepsilon}, \quad \gamma_\varepsilon^{(2)} = \gamma_\varepsilon - \gamma_\varepsilon^{(1)}. \quad (236)$$

Then

$$\sqrt{\frac{\alpha_\varepsilon}{\varepsilon}} \frac{\gamma_\varepsilon^{(1)}}{(1 + \frac{\delta_\varepsilon}{3} \gamma_+ g_\varepsilon)^{1/2}} \quad \text{is bounded in } L^2_{\text{loc}}(dt; L^2(M|v \cdot n(x)| dv d\sigma_x)), \quad (237)$$

$$\sqrt{\frac{\alpha_\varepsilon}{\varepsilon}} \frac{\gamma_\varepsilon^{(1)}}{(1 + \frac{\delta_\varepsilon}{3} \langle \gamma_+ g_\varepsilon \rangle_{\partial\Omega})^{1/2}} \\ \text{is bounded in } L^2_{\text{loc}}(dt; L^2(M|v \cdot n(x)| dv d\sigma_x)), \quad (238)$$

$$\frac{\alpha_\varepsilon}{\varepsilon \delta_\varepsilon} \gamma_\varepsilon^{(2)} \quad \text{is bounded in } L^1_{\text{loc}}(dt; L^1(M|v \cdot n(x)| dv d\sigma_x)). \quad (239)$$

5.5. Convergence toward the Euler system

We present here a method of proof based on an energy method or more precisely the relative entropy method (see [174]). Indeed contrary to the two preceding cases, we suppose here the existence of a strong solution to the Euler system and we show the convergence toward this solution. The technique used is based on a Grönwall lemma. In [115] (in collaboration with Lions), we show this convergence with an assumption on high velocities (A2). This assumption was removed in [149]. We will present the result of [149]. We introduce a defect measure (as in the Stokes case) which disappears at the limit. We take well prepared initial data (i.e., there are no acoustic waves) and the temperature fluctuation is equal to 0 initially.

5.5.1. Entropic convergence. In addition to the assumptions on G_ε^0 which we imposed in the case of convergence toward the Navier–Stokes system, we suppose that g_ε^0 converges entropically toward g^0 and that $g^0 = u^0 \cdot v$ (with $\text{div } u^0 = 0$), i.e., that

$$g_\varepsilon^0 \rightarrow g^0 \quad \text{in } w - L^1(M dv dx) \quad (240)$$

and

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \int_{\Omega} \langle h(\varepsilon g_\varepsilon^0) \rangle dx = \frac{1}{2} \int_{\Omega} \langle (g^0)^2 \rangle dx. \quad (241)$$

It is also supposed that u^0 is regular enough (for example $u^0 \in H^s$, $s > D/2 + 1$) to be able to build a strong solution $\tilde{u}$ of the Euler system with the initial data u^0 . Then, we have $\tilde{u} \in L^\infty_{\text{loc}}([0, T^*]; H^s)$ for some $T^* > 0$.

5.5.2. Relative entropy. We want to show that the distribution F_ε is close to a Maxwellian $M_{(0,\varepsilon\tilde{u},0)} = M\tilde{G}_\varepsilon$. But as F_ε is only in $L\log L$, we have to estimate the difference between F_ε and $M_{(0,\varepsilon\tilde{u},0)}$ using the relative entropy

$$H(G_\varepsilon, \tilde{G}_\varepsilon) = \int_{\Omega} \left\langle G_\varepsilon \log \left(\frac{G_\varepsilon}{\tilde{G}_\varepsilon} \right) - G_\varepsilon + \tilde{G}_\varepsilon \right\rangle. \quad (242)$$

Using the improved entropy inequality (212), we get

$$\begin{aligned} H(G_\varepsilon, \tilde{G}_\varepsilon) + \varepsilon \int_{\Omega} \text{tr}(m_\varepsilon) + \int_0^t \text{ds} D(G_\varepsilon) \\ \leq H(G_\varepsilon^0, \tilde{G}_\varepsilon^0) + \int_0^t \int_{\Omega} \langle G_\varepsilon \partial_t \log \tilde{G}_\varepsilon \rangle + \varepsilon^2 \partial_t \langle g_\varepsilon v \rangle \cdot \tilde{u} + \varepsilon^3 \partial_t \langle g_\varepsilon \rangle \frac{|\tilde{u}|^2}{2} \text{ds}, \end{aligned}$$

where m_ε denotes the sequence of defect measures appearing in the conservation of momentum.

THEOREM 5.9. *Under some assumption of the collision kernel, if G_ε is a sequence of renormalized solutions of the Boltzmann equations with initial condition G_ε^0 , and such that g_ε^0 converges entropically to $g^0 = u^0 \cdot v$, where $u^0 \in H^s$, $s > D/2 + 1$. Then, for all $0 \leq t < T^*$,*

$$g_\varepsilon(t) \rightarrow \tilde{u}(t) \cdot v \quad \text{entropically}, \quad (243)$$

where $\tilde{u}(t)$ is the unique solution of the Euler system in $L_{\text{loc}}^\infty([0, T^*]; H^s)$ with the initial condition u^0 . Moreover, the convergence is locally uniform in time.

Let us explain here the idea of the proof of the above result. It is based on a Grönwall lemma. Indeed, after some nontrivial computations, one can rewrite the entropy inequality as follows

$$\begin{aligned} \frac{1}{\varepsilon^2} \left[H(G_\varepsilon, \tilde{G}_\varepsilon) + \varepsilon \int_{\Omega} \text{tr}(m_\varepsilon) \right] (t) + \frac{1}{\varepsilon^2} \int_0^t \text{ds} D(G_\varepsilon) \\ \leq \frac{1}{\varepsilon^2} H(G_\varepsilon^0, \tilde{G}_\varepsilon^0) + \int_0^t \|\nabla \tilde{u}\|_{L^\infty} \frac{1}{\varepsilon^2} \left[H(G_\varepsilon, \tilde{G}_\varepsilon) + \varepsilon \int_{\Omega} \text{tr}(m_\varepsilon) \right] (s) \text{ds} + A_\varepsilon, \end{aligned}$$

where A_ε converges to 0. Hence, we deduce that $H(G_\varepsilon, \tilde{G}_\varepsilon)$ goes to 0 in $L_{\text{loc}}^\infty([0, T^*])$.

We want to point out that the same type of argument can be used to prove the convergence toward the Navier–Stokes system in the case a regular solution is known to exist.

6. Some homogenization problems

In this section we would like to present some homogenization problems. We will only consider examples which are related to fluid mechanics.

The homogenization of the Stokes and of the incompressible Navier–Stokes equations in a porous medium (open set perforated with tiny holes) has been studied in many works from the formal point of view as well as the rigorous one. We refer the interested reader to [21,107,152] for some formal developments and to [4,135,160] for some rigorous mathematical results.

Let us start by giving a definition of a porous medium. Let Ω be a smooth bounded domain of $\mathbb{R}^N$ and define $\mathcal{Y} =]0, 1[^N$ to be the unit open cube of $\mathbb{R}^N$. Let $\mathcal{Y}_s$ (the solid part) be a closed smooth subset of $\mathcal{Y}$ with a strictly positive measure. The fluid part is then given by $\mathcal{Y}_f = \mathcal{Y} - \mathcal{Y}_s$ and we define $\theta = |\mathcal{Y}_f|$ the Lebesgue measure of $\mathcal{Y}_f$ and we assume that $0 < \theta < 1$. The constant θ is called the porosity of the porous medium. Repeating the domain $\mathcal{Y}_f$ by $\mathcal{Y}$ -periodicity we get the fluid domain E_f which can also be defined as

$$E_f = \{y \in \mathbb{R}^N \mid \exists k \in \mathbb{Z}^N, \text{ such that } y - k \in \mathcal{Y}_f\}. \quad (244)$$

In the same way, we can define $E_s = \mathbb{R}^N - E_f$,

$$E_s = \{y \in \mathbb{R}^N \mid \exists k \in \mathbb{Z}^N, \text{ such that } y - k \in \mathcal{Y}_s\}. \quad (245)$$

It is easy to see that E_f is a connected domain, while E_s is formed by separate smooth subsets. In the sequel, we denote for all $k \in \mathbb{Z}^N$, $\mathcal{Y}^k = \mathcal{Y} + k$ the translate of the cell $\mathcal{Y}$ by the vector k , we also denote $\mathcal{Y}_s^k = \mathcal{Y}_s + k$ and $\mathcal{Y}_f^k = \mathcal{Y}_f + k$. Hence, for all ε , we can define the domain Ω_ε as the intersection of Ω with the fluid domain rescaled by ε , namely $\Omega_\varepsilon = \Omega \cap \varepsilon E_f$. However, to get a smooth connected domain, we will not remove the solid parts of the cells which intersect the boundary of Ω . We define

$$\Omega_\varepsilon = \Omega - U \{\varepsilon \mathcal{Y}_s^k, \text{ where } k \in \mathbb{Z}^N, \varepsilon \mathcal{Y}^k \subset \Omega\}.$$

We also denote $K_\varepsilon = \{k \mid k \in \mathbb{Z}^N \text{ and } \varepsilon \mathcal{Y}^k \subset \Omega\}$.

REMARK 6.1. We can also consider more general domains, especially the more physical case where E_s is a connected set of $\mathbb{R}^N$ which can be achieved by allowing $\mathcal{Y}_s$ to be a closed subset of $\overline{\mathcal{Y}}$ (this is not possible in $N = 2$ since we also want that Ω_ε is connected). We refer the interested reader to the paper of Allaire [4] where the so-called “oscillating test function” method of Tartar is extended to the case of a connected E_s .

Due to the presence of the holes $\varepsilon \mathcal{Y}_s^k$, the domain Ω_ε depends on ε and hence to study the convergence of a sequence of functions, we have to extend the functions defined in Ω_ε to the whole domain Ω . This can be done in two different ways.

DEFINITION 6.2. For any function $\phi \in L^1(\Omega_\varepsilon)$, we define

$$\tilde{\phi} = \begin{cases} \phi & \text{in } \Omega_\varepsilon, \\ 0 & \text{in } \Omega - \Omega_\varepsilon, \end{cases} \quad (246)$$

the extension by 0 of ϕ and

$$\hat{\phi} = \begin{cases} \phi & \text{in } \Omega_\varepsilon, \\ \frac{1}{\varepsilon|\mathcal{Y}_f|} \int_{\varepsilon\mathcal{Y}_f^k} \phi \, dy & \text{in } \varepsilon\mathcal{Y}_s^k \end{cases} \quad \forall k \in K_\varepsilon. \quad (247)$$

We will also need the restriction operator constructed by Tartar [160] for the case of a solid part $\mathcal{Y}_s$ strictly included in $\mathcal{Y}$ and by Allaire [4] for more general conditions on the solid part.

LEMMA 6.3. *There exists a linear operator R_ε from $H_0^1(\Omega)^N$ to $H_0^1(\Omega_\varepsilon)^N$ (called restriction operator) such that*

- (i) $\forall \phi \in H_0^1(\Omega_\varepsilon)^N$, we have $R_\varepsilon \tilde{\phi} = \phi$;
- (ii) $\nabla \cdot u = 0$ in Ω implies that $\nabla \cdot R_\varepsilon u = 0$ in Ω_ε ;
- (iii) there exists a constant C such that for all $u \in H_0^1(\Omega)^N$, we have

$$\|R_\varepsilon u\|_{L^2(\Omega_\varepsilon)} + \varepsilon \|\nabla(R_\varepsilon u)\|_{L^2(\Omega_\varepsilon)} \leq C[\|u\|_{L^2(\Omega)} + \varepsilon \|\nabla u\|_{L^2(\Omega)}]. \quad (248)$$

The operator R_ε defined above also acts from $W_0^{1,r}(\Omega)$ into $W_0^{1,r}(\Omega_\varepsilon)$ for all $1 < r < \infty$ and we have an estimate similar to (248) where the L^2 norms are replaced by L^r norms.

Due to the presence of the holes in the domain Ω_ε , the Poincaré's inequality reads the lemma.

LEMMA 6.4. *There exists a constant C which depends only on $\mathcal{Y}_s$ such that for all $u \in W_0^{1,p}(\Omega_\varepsilon)$, we have*

$$\|u\|_{L^p(\Omega_\varepsilon)} \leq C\varepsilon \|\nabla u\|_{L^p(\Omega_\varepsilon)}. \quad (249)$$

We refer to [160] for a proof of this lemma. By a simple duality argument we also have the following relation for all $1 < p < \infty$,

$$\|u\|_{W^{-1,p}(\Omega_\varepsilon)} \leq C\varepsilon \|u\|_{L^p(\Omega_\varepsilon)}. \quad (250)$$

Finally, we define the permeability matrix $\bar{A}$. For all i , $1 \leq i \leq N$, let $(v_i, q_i) \in H^1(\mathcal{Y}_f)^N \times L^2(\mathcal{Y}_f)/\mathbb{R}$ be the unique solution of the following system

$$\begin{cases} -\Delta v_i + \nabla q_i = e_i & \text{in } \mathcal{Y}_f, \\ \operatorname{div} v_i = 0 & \text{in } \mathcal{Y}_f, \\ v_i = 0 & \text{on } \partial\mathcal{Y}_s, \text{ and } v_i, q_i \text{ are } \mathcal{Y}\text{-periodic.} \end{cases} \quad (S_i)$$

Using regularity results of the Stokes problem, we infer that v_i and q_i are smooth. We extend v_i to the whole domain $\mathcal{Y}$ by setting $v_i(y) = 0$ if $y \in \mathcal{Y}_s$. Then, for all $y \in \mathcal{Y}_f$, $A(y)$ is taken to be the matrix composed of the column vectors $v_i(y)$ and $\bar{A} = \int_{\mathcal{Y}_f} A(y) \, dy$.

It is easy to see that $\bar{A}$ is a symmetric positive definite matrix. Indeed, multiplying the first equation in (S_i) by v_j and the first equation in (S_i) by v_i , we get that $\int_{\mathcal{Y}_f} \nabla v_i \cdot \nabla v_j =$

$\int_{\mathcal{Y}_f} v_{ji} = \bar{A}_{ji}$ and $\int_{\mathcal{Y}_f} \nabla v_j \cdot \nabla v_i = \int_{\mathcal{Y}_f} v_{ij} = \bar{A}_{ij}$ where we wrote $v_i(y) = \sum_{j=1}^N v_{ji}(y)e_j$. Then to prove that $\bar{A}$ is positive definite, we just notice that for all vector $X = \sum_{j=1}^N x_j e_j$, we have $\sum_{ij} x_i \bar{A}_{ij} x_j = \|\nabla \sum_{j=1}^N x_j v_j\|_{L^2(\mathcal{Y}_f)}^2$ and that $\{v_i, 1 \leq i \leq N\}$ is an independent family.

6.1. Darcy law

Let us start by recalling the derivation of the Darcy law [48]. We consider the Stokes problem in the domain Ω_ε ,

$$\begin{cases} -\Delta u_\varepsilon + \nabla p_\varepsilon = f, \\ \operatorname{div} u_\varepsilon = 0, \quad u_\varepsilon = 0 \text{ on } \partial\Omega_\varepsilon. \end{cases} \quad (252)$$

THEOREM 6.5. *Prolonging u_ε by zero in the holes, we have the following convergence*

$$\tilde{u}_\varepsilon \rightarrow u \quad \text{weakly in } (L^2(\Omega)), \quad (253)$$

where $u = \bar{A}(f - \nabla p)$ and satisfies $\operatorname{div} u = 0$. This is the Darcy law.

The proof uses the “oscillating test function” method of Tartar [160]. Indeed, testing (252) with $\phi(x)v_i(x/\varepsilon)$ where $\phi \in C_0^\infty(\Omega)$, we can pass to the weakly to the limit in the different terms to deduce (253). Actually some nontrivial work should be done to pass to the limit in the pressure term and we refer to [160] and [4].

6.2. Homogenization of a compressible model

Here, we give a derivation of the porous medium equation. We start with the following semistationary model

$$\begin{cases} \varepsilon^2 \partial_t \rho_\varepsilon + \operatorname{div}(\rho_\varepsilon u_\varepsilon) = 0, \\ -\mu \Delta u_\varepsilon - \xi \nabla \operatorname{div} u_\varepsilon + \nabla \rho_\varepsilon^\gamma = \rho_\varepsilon f + g \end{cases} \quad (254)$$

complemented with the boundary condition $u_\varepsilon = 0$ on $\partial\Omega_\varepsilon$ and the initial condition $\rho_\varepsilon(t=0) = \rho_{\varepsilon 0}$. The force term is such that $f \in L^\infty((0, T) \times \Omega_\varepsilon)$ and $g \in L^2((0, T) \times \Omega_\varepsilon)$. We also assume that $\gamma \geq 1$ and that $\|f\|_{L^\infty}$ is small enough if $\gamma = 1$.

We assume that the initial data is such that $\rho_{\varepsilon 0} \in L^1 \cap L^\gamma(\Omega_\varepsilon)$ if $\gamma > 1$, that $\int_{\Omega_\varepsilon} \rho_{\varepsilon 0} |\log \rho_{\varepsilon 0}| < C$ if $\gamma = 1$ and that $\hat{\rho}_{\varepsilon 0}$ converges weakly to ρ_0 in $L^\gamma(\Omega)$.

We consider a sequence of weak solutions $(\rho_\varepsilon, u_\varepsilon)$ of the semistationary model (254) such that for all $T > 0$, $\rho_\varepsilon \in C([0, T]; L^1(\Omega_\varepsilon)) \cap L^\infty(0, T; L^\gamma(\Omega_\varepsilon)) \cap L^{2\gamma}((0, T) \times \Omega_\varepsilon)$ and $\rho_\varepsilon |\log \rho_\varepsilon| \in L^\infty(0, T; L^1(\Omega_\varepsilon))$ if $\gamma = 1$. Moreover, u_ε is such that $u_\varepsilon/\varepsilon \in L^2(0, T; H_0^1(\Omega_\varepsilon))$ and $u_\varepsilon/\varepsilon^2 \in L^2((0, T) \times \Omega_\varepsilon)$. Finally, we also require that $\hat{\rho}_\varepsilon$ is

bounded in $L_T^2(H^1(\Omega)) + \varepsilon L_T^2(L^2(\Omega))$. We assume that the bounds given above are uniform in ε . We point out that the fact that we can consider a sequence of solutions satisfying the above uniform estimates can be proved using the methods of [109].

Before studying the limit of the sequence $(u_\varepsilon, \rho_\varepsilon, p_\varepsilon)$, we have to prolong it to Ω . Let $\tilde{u}_\varepsilon$, $\tilde{\rho}_\varepsilon$ and $\hat{p}_\varepsilon$ be the extensions of u_ε , ρ_ε and p_ε to the whole domain Ω .

THEOREM 6.6. *Under the above assumptions,*

$$\begin{aligned}\tilde{\rho}_\varepsilon &\rightarrow \theta \rho \quad \text{weakly in } L_T^r(L^\gamma(\Omega)) \cap L^{2\gamma}((0, T) \times \Omega), \\ \hat{\rho}_\varepsilon &\rightarrow \rho \quad \text{strongly in } L_T^r(L^\gamma(\Omega)) \cap L^{\gamma+1}((0, T) \times \Omega), \\ \frac{\tilde{u}_\varepsilon}{\varepsilon^2} &\rightarrow u \quad \text{weakly in } L_T^2(L^2(\Omega))\end{aligned}$$

for all $r < \infty$, where $\rho \in L^{2\gamma}((0, T) \times \Omega)$, $\rho^\gamma \in L_T^2(H^1(\Omega))$ and ρ is the solution of the following system

$$\begin{cases} \theta \partial_t \rho + \frac{1}{\mu} \operatorname{div} [\rho \bar{A}(\rho f + g - \nabla \rho^\gamma)] = 0, \\ \rho \bar{A}(\rho f + g - \nabla \rho^\gamma) \cdot n = 0 \quad \text{on } \partial\Omega, \\ \rho(t=0) = \rho_0, \end{cases} \quad (255)$$

and u is given by

$$u = \bar{A}(\rho f + g - \nabla \rho^\gamma) \quad \text{on } \{\rho > 0\}. \quad (256)$$

We point out here that even though each one of the terms f , g and $\nabla \rho^\gamma$ does not have necessary a trace on the boundary $\partial\Omega$, the combination of them appearing in (255) has a sense. A formal derivation of the system (255) can be found in [55]. The relation (256) giving u as a function of the pressure is a Darcy law [48,160].

REMARK 6.7. If $\bar{A} = \alpha I$ (which is the case if for instance $\mathcal{Y}_s$ is a ball) and $f = g = 0$ then we get the following system

$$\begin{cases} \partial_t \rho - \beta \Delta \rho^{\gamma+1} = 0, \\ \frac{\partial \rho^{\gamma+1}}{\partial n} = 0 \quad \text{on } \partial\Omega, \\ \rho(t=0) = \rho_0, \end{cases} \quad (257)$$

where $\beta = \frac{\alpha\gamma}{\theta\mu(\gamma+1)}$. This system is the so-called ‘‘porous medium’’ equation.

6.3. Homogenization of the Euler system

We consider an incompressible perfect fluid governed by the Euler equation. We consider the following system of equations

$$\begin{cases} \partial_t u^\varepsilon + \varepsilon u^\varepsilon \cdot \nabla u^\varepsilon = -\nabla p^\varepsilon + f^\varepsilon(x), \\ \operatorname{div}(u^\varepsilon) = 0, \\ u^\varepsilon \cdot n = 0 \quad \text{on } \partial\Omega_\varepsilon, \\ u^\varepsilon|_{t=0} = u_0^\varepsilon, \end{cases} \quad (258)$$

where u^ε is the velocity, p^ε is the pressure, f^ε is an exterior force and n is the outward normal vector to Ω_ε . Arguing as in the book of Bensoussan, Lions and Papanicolaou [21] (see also [107]) and the book of Sanchez-Palencia [152], we make an asymptotic development using both a microscopic scale and a macroscopic scale. Hence, we can derive a (formal) limit system. Indeed taking u^ε of the form $u^\varepsilon = u^0(t, x, x/\varepsilon) + \varepsilon u^1(t, x, x/\varepsilon) + \dots$, we get formally the following system, for $v(t, x, y) = u^0(t, x, y)$,

$$\begin{cases} \partial_t v + v \cdot \nabla_y v = -\nabla_y p(x, y) - \nabla_x q(x) + f(t, x, y), \\ \operatorname{div}_y(v) = 0, \quad \operatorname{div}_x(\int_{\mathcal{Y}_f} v(x, y) dy) = 0, \\ v(x, y) \cdot n = 0 \quad \text{on } \Omega \times \partial\mathcal{Y}_s, \\ (\int_{\mathcal{Y}_f} v(x, y) dy) \cdot n = 0 \quad \text{on } \partial\Omega, \\ v|_{t=0} = v_0, \end{cases} \quad (259)$$

where $f(t, x, y)$ and $v_0(x, y)$ are the two-scale limits of the sequences f^ε and u_0^ε and here n is the inward normal vector to $\mathcal{Y}_s$. The notion of two-scale convergence is aimed at a better description of sequences of oscillating functions with a known scale. It was introduced by Nguetseng [137,138] and later extended by Allaire [5] where one can find the mathematical setting we use here.

DEFINITION 6.8. Let u^ε be a sequence of functions such that $u^\varepsilon \in L^2(\Omega_\varepsilon)$ and $\|u^\varepsilon\|_{L^2(\Omega_\varepsilon)}$ is bounded uniformly in ε . If $v(x, y) \in L^2(\Omega \times \mathcal{Y}_f)$, then we say that u^ε two-scale converges to v if and only if $\forall \psi \in C(\Omega \times \mathcal{Y}_f)$, we have

$$\lim_{\varepsilon \rightarrow 0} \int_{\Omega_\varepsilon} u^\varepsilon(x) \psi\left(x, \frac{x}{\varepsilon}\right) dx = \int_{\Omega \times \mathcal{Y}_f} v(x, y) \psi(x, y) dx dy. \quad (260)$$

Moreover, we say that u^ε two-scale converges strongly to v if and only if $v(x, y) \in L^2(\Omega, C(\mathcal{Y}_f))$ and we have

$$\lim_{\varepsilon \rightarrow 0} \left\| u^\varepsilon(x) - v\left(x, \frac{x}{\varepsilon}\right) \right\|_{L^2(\tilde{\Omega}_\varepsilon)} = 0 \quad (261)$$

and

$$\lim_{\varepsilon \rightarrow 0} \|u^\varepsilon(x)\|_{L^2(\Omega_\varepsilon - \tilde{\Omega}_\varepsilon)} = 0. \quad (262)$$

We will state two results. The first one concerns the Cauchy problem for the limit system and the second one concerns the convergence of a sequence of the solutions to (258) toward a solution to (259). We start by defining the following functional spaces

$$\mathcal{A} = \{v(x, y), v \in L^2(\Omega \times \mathcal{Y}_f), \operatorname{div}_y(v) = 0, \operatorname{div}_x(\bar{v}) = 0, \\ v \cdot n = 0 \text{ on } \Omega \times \partial\mathcal{Y}_s, \bar{v} \cdot n = 0 \text{ on } \partial\Omega\} \quad (263)$$

$$\mathcal{A}_\infty = \{v(x, y), v \in \mathcal{A} \text{ and } \operatorname{curl}_y(v) \in L^\infty(\Omega \times \mathcal{Y}_f)\}, \quad (264)$$

where div_y and div_x denote respectively the divergence in the y and in the x variables, namely $\operatorname{div}_y(v) = \partial_{y_1}v_1 + \partial_{y_2}v_2$ and $\operatorname{div}_x(v) = \partial_{x_1}v_1 + \partial_{x_2}v_2$. Moreover, $\bar{v}$ denotes the integral of v over $\mathcal{Y}_f$, namely $\bar{v}(x) = \int_{\mathcal{Y}_f} v(x, y) dy$. Finally, n denotes the exterior normal vector to $\partial\mathcal{Y}_f$ or to $\partial\Omega$.

Now, we give an existence result for the limit system (259).

THEOREM 6.9. *Take $v_0 \in \mathcal{A}_\infty$ and $f \in L^1((0, \infty); \mathcal{A}_\infty)$. Then, there exists a global solution to the system (259) such that*

$$v \in C([0, \infty); \mathcal{A}) \cap L^\infty((0, \infty); \mathcal{A}_\infty). \quad (265)$$

This result is similar to the existence result for the incompressible Euler system by Yudovich [175]. However, unlike Yudovich solutions, the uniqueness of the solutions constructed in Theorem 6.9 is not known.

Now, we focus on the convergence result. We have to assume that u_0^ε is bounded in $L^3(\Omega_\varepsilon)$, $\operatorname{div}(u_0^\varepsilon) = 0$, $u_0^\varepsilon \cdot n = 0$ on $\partial\Omega_\varepsilon$, $\varepsilon \operatorname{curl}(u_0^\varepsilon)$ is in L^∞ (which implies the existence and uniqueness for the initial system) and that u_0^ε two-scale converges strongly to v_0 where $v_0 \in \mathcal{A}_\infty$. Moreover, we assume that f^ε is divergence-free, that it is bounded in $L^1((0, \infty); L^3(\Omega_\varepsilon))$, that $\operatorname{curl} f^\varepsilon$ is bounded in $L^1((0, \infty); L^\infty(\Omega_\varepsilon))$ and that f^ε two-scale converges strongly to f , namely

$$\lim_{\varepsilon \rightarrow 0} \left\| u_0^\varepsilon(x) - v_0\left(x, \frac{x}{\varepsilon}\right) \right\|_{L^2(\Omega_\varepsilon)} = 0, \quad (266)$$

$$\lim_{\varepsilon \rightarrow 0} \left\| f^\varepsilon(t, x) - f\left(t, x, \frac{x}{\varepsilon}\right) \right\|_{L^1((0, \infty); L^2(\Omega_\varepsilon))} = 0, \quad (267)$$

where v_0 and f satisfy the hypotheses of Theorem 6.9. Here, we only take the two-scale convergence in the x variable, then we have the theorem.

THEOREM 6.10. *Under the above conditions there exists a sequence u^ε of solutions to the initial system (258). Moreover, extracting a subsequence if necessary u^ε two-scale converges to v where v is a solution to the limit system (259).*

We refer to [116, 128] for the proof.

7. Conclusion

Before giving some concluding remarks we would like to mention some other limit problems which we did not develop in the previous sections. These asymptotic problems are very important and we want to give some references to the interested reader.

7.1. Other limits

7.1.1. The infinite Prandtl number limit. The infinite Prandtl number limit was considered in [170] (see equation (8) for the definition of the Prandtl number). At the limit the so-called infinite Prandtl number convection system is retrieved at the limit. It is a system where the velocity is slaved by the temperature field since velocity diffuses more rapidly than the temperature. The proof is based on an expansion using two time scales.

7.1.2. The zero surface tension limit. The infinite Weber limit was considered in [7]. This is the same as the zero surface tension limit. It was proved in [7] that when surface tension goes to zero the water wave system with surface tension [6] converges to the water wave system without surface tension [172]. This is a singular limit since surface tension has a regularizing effect even though the initial system and the limit system are of the same type.

7.1.3. The quasineutral limit. The convergence from the Vlasov–Poisson system toward the incompressible Euler equation in the quasineutral limit was considered in [24] and [126]. These two works deal with the zero temperature case, namely the density $f(t, x, v)$ is a delta function in velocity.

A related problem, is the relation between the Euler system and the N vortices problem. This was considered in [121]. We also refer to [120] for an inviscid limit with concentrated vorticity.

For related asymptotic problems in plasma physics, we refer to [131] for the limit from the Klein–Gordon–Zakharov system to the nonlinear Schrödinger equation. We also refer to [130] for the limit from Maxwell–Klein–Gordon and Maxwell–Dirac to Poisson–Schrödinger when the speed of light c goes to infinity.

7.1.4. Thin domains. Fluid equations considered in thin domains give rise to many asymptotic problems (see [91] and [146] and the references therein). Indeed, taking for instance the Navier–Stokes equation in a thin domain $(0, \varepsilon) \times \mathbb{T}^2$, we can try to describe the solutions when ε goes to zero. To do so, we have to make a change of variable and rescale the domain to a fixed domain $(0, 1) \times \mathbb{T}^2$. This introduces a small parameter ε in the equation written in the fixed domain. The small parameter ε is the ratio between the vertical length scale and the horizontal one.

7.2. Concluding remarks

As can be seen from the different section of this chapter, asymptotic problem in hydrodynamics is a vast subject by the number of problems one can consider and the number of

methods used to treat them. It is an important subject from physical and numerical point of view. Besides, it is the motor behind the development of many new mathematical tools such as (the group method, defect measures, boundary layer theory...) to handle the several physical phenomenon such as (oscillations, boundary layers...).

In this review paper, we tried to give an idea about some of the advances made in these singular limits during the last few years. At several places, the author put more emphasis on results he is more aware of.

Acknowledgement

The author is partially supported by an NSF Grant DMS-04-03983.

References

- [1] H. Abidi and R. Danchin, *Optimal bounds for the inviscid limit of Navier–Stokes equations*, Asymptot. Anal. **38** (1) (2004), 35–46.
- [2] T. Alazard, *Incompressible limit of the nonisentropic Euler equations with the solid wall boundary conditions*, Adv. Differential Equations **10** (1) (2005), 19–44.
- [3] T. Alazard, *Low Mach number limit of the full Navier–Stokes equations*, Arch. Ration. Mech. Anal. **180** (1) (2006), 1–73.
- [4] G. Allaire, *Homogenization of the Stokes flow in a connected porous medium*, Asymptotic Anal. **2** (3) (1989), 203–222.
- [5] G. Allaire, *Homogenization and two-scale convergence*, SIAM J. Math. Anal. **23** (6) (1992), 1482–1518.
- [6] D.M. Ambrose, *Well-posedness of vortex sheets with surface tension*, SIAM J. Math. Anal. **35** (2003), 211–244.
- [7] D.M. Ambrose and N. Masmoudi, *The zero surface tension limit of two-dimensional water waves*, Comm. Pure Appl. Math. **58** (10) (2005), 1287–1315.
- [8] A. Babin, A. Mahalov and B. Nicolaenko, *Global splitting, integrability and regularity of 3D Euler and Navier–Stokes equations for uniformly rotating fluids*, European J. Mech. B Fluids **15** (3) (1996), 291–300.
- [9] A. Babin, A. Mahalov and B. Nicolaenko, *Regularity and integrability of 3D Euler and Navier–Stokes equations for rotating fluids*, Asymptot. Anal. **15** (2) (1997), 103–150.
- [10] A. Babin, A. Mahalov and B. Nicolaenko, *3D Navier–Stokes and Euler equations with initial data characterized by uniformly large vorticity*, Indiana Univ. Math. J. **50** (Special Issue) (2001), 1–35. Dedicated to Professors Ciprian Foias and Roger Temam (Bloomington, IN, 2000).
- [11] C. Bardos, *Existence et unicité de la solution de l'équation d'Euler en dimension deux*, J. Math. Anal. Appl. **40** (1972), 769–790.
- [12] C. Bardos, F. Golse and C.D. Levermore, *Fluid dynamic limits of kinetic equations. II. Convergence proofs for the Boltzmann equation*, Comm. Pure Appl. Math. **46** (5) (1993), 667–753.
- [13] C. Bardos, F. Golse and C.D. Levermore, *Acoustic and Stokes limits for the Boltzmann equation*, C. R. Acad. Sci. Paris Sér. I Math. **327** (3) (1998), 323–328.
- [14] C. Bardos, F. Golse and C.D. Levermore, *The acoustic limit for the Boltzmann equation*, Arch. Ration. Mech. Anal. **153** (3) (2000), 177–204.
- [15] C. Bardos, F. Golse and D. Levermore, *Fluid dynamic limits of kinetic equations. I. Formal derivations*, J. Statist. Phys. **63** (1/2) (1991), 323–344.
- [16] C. Bardos and S. Ukai, *The classical incompressible Navier–Stokes limit of the Boltzmann equation*, Math. Models Methods Appl. Sci. **1** (2) (1991), 235–257.
- [17] H. Beirão da Veiga, *An L^p -theory for the n -dimensional, stationary, compressible Navier–Stokes equations, and the incompressible limit for compressible fluids. The equilibrium solutions*, Comm. Math. Phys. **109** (2) (1987), 229–248.

- [18] H. Beirão da Veiga, *On the singular limit for slightly compressible fluids*, Calc. Var. Partial Differential Equations **2** (2) (1994), 205–218.
- [19] H. Beirão da Veiga, *Singular limits in compressible fluid dynamics*, Arch. Rational Mech. Anal. **128** (4) (1994), 313–327.
- [20] J. Benameur, S. Ibrahim and M. Majdoub, *Asymptotic study of a magneto-hydrodynamic system*, Differential Integral Equations **18** (3) (2005), 299–324.
- [21] A. Bensoussan, J.-L. Lions and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, North-Holland, Amsterdam (1978).
- [22] A.L. Bertozzi and P. Constantin, *Global regularity for vortex patches*, Comm. Math. Phys. **152** (1) (1993), 19–28.
- [23] A.J. Bourgeois and J.T. Beale, *Validity of the quasigeostrophic model for large-scale flow in the atmosphere and ocean*, SIAM J. Math. Anal. **25** (4) (1994), 1023–1068.
- [24] Y. Brenier, *Convergence of the Vlasov–Poisson system to the incompressible Euler equations*, Comm. Partial Differential Equations **25** (3/4) (2000), 737–754.
- [25] D. Bresch, B. Desjardins and D. Gérard-Varet, *Rotating fluids in a cylinder*, Discrete Contin. Dyn. Syst. **11** (1) (2004), 47–82.
- [26] D. Bresch, B. Desjardins, E. Grenier and C.-K. Lin, *Low Mach number limit of viscous polytropic flows: Formal asymptotics in the periodic case*, Stud. Appl. Math. **109** (2) (2002), 125–149.
- [27] D. Bresch, F. Guillén-González, N. Masmoudi and M.A. Rodríguez-Bellido, *Asymptotic derivation of a Navier condition for the primitive equations*, Asymptot. Anal. **33** (3/4) (2003), 237–259.
- [28] R.E. Caflisch, *The fluid dynamic limit of the nonlinear Boltzmann equation*, Comm. Pure Appl. Math. **33** (5) (1980), 651–666.
- [29] R.E. Caflisch and M. Sammartino, *Existence and singularities for the Prandtl boundary layer equations*, ZAMM Z. Angew. Math. Mech. **80** (11/12) (2000), 733–744. Special issue on the occasion of the 125th anniversary of the birth of Ludwig Prandtl.
- [30] C. Cercignani, *The Boltzmann Equation and Its Applications*, Springer-Verlag, New York (1988).
- [31] C. Cercignani, R. Illner and M. Pulvirenti, *The Mathematical Theory of Dilute Gases*, Springer-Verlag, New York (1994).
- [32] J.-Y. Chemin, *Persistence de structures géométriques dans les fluides incompressibles bidimensionnels*, Ann. Sci. École Norm. Sup. (4) **26** (4) (1993), 517–542.
- [33] J.-Y. Chemin, *Fluides parfaits incompressibles*, Astérisque **230** (1995).
- [34] J.-Y. Chemin, *A remark on the inviscid limit for two-dimensional incompressible fluids*, Comm. Partial Differential Equations **21** (11/12) (1996), 1771–1779.
- [35] J.-Y. Chemin, B. Desjardins, I. Gallagher and E. Grenier, *Ekman boundary layers in rotating fluids*, ESAIM Control Optim. Calc. Var. **8** (electronic) (2002), 441–466. A tribute to J.L. Lions.
- [36] T. Clopeau, A. Mikelić and R. Robert, *On the vanishing viscosity limit for the 2D incompressible Navier–Stokes equations with the friction type boundary conditions*, Nonlinearity **11** (6) (1998), 1625–1636.
- [37] T. Colin and P. Fabrie, *Rotating fluid at high Rossby number driven by a surface stress: Existence and convergence*, Adv. Differential Equations **2** (5) (1997), 715–751.
- [38] P. Constantin, *Note on loss of regularity for solutions of the 3-D incompressible Euler and related equations*, Comm. Math. Phys. **104** (2) (1986), 311–326.
- [39] P. Constantin and C. Foias, *Navier–Stokes Equations*, University of Chicago Press, Chicago, IL (1988).
- [40] P. Constantin and J. Wu, *Inviscid limit for vortex patches*, Nonlinearity **8** (5) (1995), 735–742.
- [41] P. Constantin and J. Wu, *The inviscid limit for non-smooth vorticity*, Indiana Univ. Math. J. **45** (1) (1996), 67–81.
- [42] R. Danchin, *Poches de tourbillon visqueuses*, J. Math. Pures Appl. (9) **76** (7) (1997), 609–647.
- [43] R. Danchin, *Persistence de structures géométriques et limite non visqueuse pour les fluides incompressibles en dimension quelconque*, Bull. Soc. Math. France **127** (2) (1999), 179–227.
- [44] R. Danchin, *Global existence in critical spaces for compressible Navier–Stokes equations*, Invent. Math. **141** (3) (2000), 579–614.
- [45] R. Danchin, *Zero Mach number limit for compressible flows with periodic boundary conditions*, Amer. J. Math. **124** (6) (2002), 1153–1219.
- [46] R. Danchin, *Zero Mach number limit in critical spaces for compressible Navier–Stokes equations*, Ann. Sci. École Norm. Sup. (4) **35** (1) (2002), 27–75.

- [47] R. Danchin, *Low Mach number limit for viscous compressible flows*, M2AN Math. Model. Numer. Anal. **39** (3) (2005), 459–475.
- [48] H. Darcy, *Les fontaines publiques de la ville de Dijon*, Dalmont, Paris (1856).
- [49] A. De Masi, R. Esposito and J.L. Lebowitz, *Incompressible Navier–Stokes and Euler limits of the Boltzmann equation*, Comm. Pure Appl. Math. **42** (8) (1989), 1189–1214.
- [50] J.-M. Delort, *Existence de nappes de tourbillon en dimension deux*, J. Amer. Math. Soc. **4** (3) (1991), 553–586.
- [51] B. Desjardins, E. Dormy and E. Grenier, *Stability of mixed Ekman–Hartmann boundary layers*, Nonlinearity **12** (2) (1999), 181–199.
- [52] B. Desjardins and E. Grenier, *Low Mach number limit of viscous compressible flows in the whole space*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **455** (1986) (1999), 2271–2279.
- [53] B. Desjardins and E. Grenier, *Linear instability implies nonlinear instability for various types of viscous boundary layers*, Ann. Inst. H. Poincaré Anal. Non Linéaire **20** (1) (2003), 87–106.
- [54] B. Desjardins, E. Grenier, P.-L. Lions and N. Masmoudi, *Incompressible limit for solutions of the isentropic Navier–Stokes equations with Dirichlet boundary conditions*, J. Math. Pures Appl. (9) **78** (5) (1999), 461–471.
- [55] J.I. Díaz, *Two problems in homogenization of porous media*, Proceedings of the Second International Seminar on Geometry, Continua and Microstructure (Getafe, 1998), Vol. 14 (1999), 141–155.
- [56] R.J. DiPerna and P.-L. Lions, *On the Cauchy problem for Boltzmann equations: Global existence and weak stability*, Ann. of Math. (2) **130** (2) (1989), 321–366.
- [57] R.J. DiPerna and A. Majda, *Reduced Hausdorff dimension and concentration–cancellation for two-dimensional incompressible flow*, J. Amer. Math. Soc. **1** (1) (1988), 59–95.
- [58] R.J. DiPerna and A.J. Majda, *Concentrations in regularizations for 2-D incompressible flow*, Comm. Pure Appl. Math. **40** (3) (1987), 301–345.
- [59] R.J. DiPerna and A.J. Majda, *Oscillations and concentrations in weak solutions of the incompressible fluid equations*, Comm. Math. Phys. **108** (4) (1987), 667–689.
- [60] W. E, *Boundary layer theory and the zero-viscosity limit of the Navier–Stokes equation*, Acta Math. Sin. (Engl. Ser.) **16** (2) (2000), 207–218.
- [61] W. E and B. Engquist, *Blowup of solutions of the unsteady Prandtl’s equation*, Comm. Pure Appl. Math. **50** (12) (1997), 1287–1293.
- [62] D.G. Ebin, *The motion of slightly compressible fluids viewed as a motion with strong constraining force*, Ann. of Math. (2) **105** (1) (1977), 141–200.
- [63] P.F. Embid and A.J. Majda, *Averaging over fast gravity waves for geophysical flows with arbitrary potential vorticity*, Comm. Partial Differential Equations **21** (3/4) (1996), 619–658.
- [64] E. Feireisl, *On compactness of solutions to the compressible isentropic Navier–Stokes equations when the density is not square integrable*, Comment. Math. Univ. Carolin. **42** (1) (2001), 83–98.
- [65] E. Feireisl, *Viscous and/or heat conducting compressible fluids*, Handbook of Mathematical Fluid Dynamics, Vol. I, North-Holland, Amsterdam (2002), 307–371.
- [66] E. Feireisl, *Dynamics of Viscous Compressible Fluids*, Oxford Lecture Series in Mathematics and Its Applications, Vol. 26, Oxford University Press, Oxford (2004).
- [67] E. Feireisl and H. Petzeltová, *On integrability up to the boundary of the weak solutions of the Navier–Stokes equations of compressible flow*, Comm. Partial Differential Equations **25** (3/4) (2000), 755–767.
- [68] I. Gallagher, *Applications of Schochet’s methods to parabolic equations*, J. Math. Pures Appl. (9) **77** (10) (1998), 989–1054.
- [69] I. Gallagher, *A remark on smooth solutions of the weakly compressible periodic Navier–Stokes equations*, J. Math. Kyoto Univ. **40** (3) (2000), 525–540.
- [70] I. Gallagher, *Résultats récents sur la limite incompressible*, Séminaire Bourbaki **926** (2004).
- [71] N. Garofalo and F. Segàla, *Another step toward the solution of the Pompeiu problem in the plane*, Comm. Partial Differential Equations **18** (3/4) (1993), 491–503.
- [72] P. Gérard, *Microlocal defect measures*, Comm. Partial Differential Equations **16** (11) (1991), 1761–1794.
- [73] D. Gérard-Varet, *Highly rotating fluids in rough domains*, J. Math. Pures Appl. (9) **82** (11) (2003), 1453–1498.
- [74] F. Golse and C.D. Levermore, *Stokes–Fourier and acoustic limits for the Boltzmann equation: Convergence proofs*, Comm. Pure Appl. Math. **55** (3) (2002), 336–393.

- [75] F. Golse, P.-L. Lions, B. Perthame and R. Sentis, *Regularity of the moments of the solution of a transport equation*, J. Funct. Anal. **76** (1) (1988), 110–125.
- [76] F. Golse, B. Perthame and R. Sentis, *Un résultat de compacité pour les équations de transport et application au calcul de la limite de la valeur propre principale d'un opérateur de transport*, C. R. Acad. Sci. Paris Sér. I Math. **301** (7) (1985), 341–344.
- [77] F. Golse and L. Saint-Raymond, *Navier–Stokes–Fourier limit for the Boltzmann equation: Convergence proofs*, Preprint (2002).
- [78] F. Golse and L. Saint-Raymond, *Velocity averaging in L^1 for the transport equation*, C. R. Math. Acad. Sci. Paris **334** (7) (2002), 557–562.
- [79] F. Golse and L. Saint-Raymond, *The Navier–Stokes limit of the Boltzmann equation for bounded collision kernels*, Invent. Math. **155** (1) (2004), 81–161.
- [80] H. Greenspan, *The Theory of Rotating Fluids*, Cambridge Monographs on Mechanics and Applied Mathematics (1969).
- [81] E. Grenier, *Oscillatory perturbations of the Navier–Stokes equations*, J. Math. Pures Appl. (9) **76** (6) (1997), 477–498.
- [82] E. Grenier, *On the nonlinear instability of Euler and Prandtl equations*, Comm. Pure Appl. Math. **53** (9) (2000), 1067–1091.
- [83] E. Grenier, *Boundary layers*, Handbook of Mathematical Fluid Dynamics, Vol. III, North-Holland, Amsterdam (2004), 245–309.
- [84] E. Grenier and N. Masmoudi, *Ekman layers of rotating fluids, the case of well prepared initial data*, Comm. Partial Differential Equations **22** (5/6) (1997), 953–975.
- [85] J.-P. Guiraud and J.-S. Darrozès, *Généralisation formelle du théorème H en présence de parois*, C. R. Acad. Sci. Paris **262** (1966), 368–371.
- [86] Y. Guo, *The diffusive limit in the Boltzmann theory beyond the Navier–Stokes approximation*, Preprint (2005).
- [87] T. Hagstrom and J. Lorenz, *All-time existence of classical solutions for slightly compressible flows*, SIAM J. Math. Anal. **29** (3) (1998), 652–672 (electronic).
- [88] D. Hilbert, *Begründung der kinetischen Gastheorie*, Math. Annalen **72** (1912), 562–577.
- [89] D. Hoff, *The zero-Mach limit of compressible flows*, Comm. Math. Phys. **192** (3) (1998), 543–554.
- [90] E. Hopf, *Über die Anfangswertaufgabe für die hydrodynamischen Grundgleichungen*, Math. Nachr. **4** (1951), 213–231.
- [91] D. Iftimie and G. Raugel, *Some results on the Navier–Stokes equations in thin 3D domains*, J. Differential Equations **169** (2) (2001), 281–331. Special issue in celebration of Jack K. Hale's 70th birthday, Part 4 (Atlanta, GA/Lisbon, 1998).
- [92] H. Isozaki, *Singular limits for the compressible Euler equation in an exterior domain*, J. Reine Angew. Math. **381** (1987), 1–36.
- [93] H. Isozaki, *Wave operators and the incompressible limit of the compressible Euler equation*, Comm. Math. Phys. **110** (3) (1987), 519–524.
- [94] J.-L. Joly, G. Métivier and J. Rauch, *Coherent and focusing multidimensional nonlinear geometric optics*, Ann. Sci. École Norm. Sup. (4) **28** (1) (1995), 51–113.
- [95] T. Kato, *Nonstationary flows of viscous and ideal fluids in R^3* , J. Functional Analysis **9** (1972), 296–305.
- [96] T. Kato, *Remarks on zero viscosity limit for nonstationary Navier–Stokes flows with boundary*, Seminar on Nonlinear Partial Differential Equations (Berkeley, CA, 1983), Springer-Verlag, New York (1984), 85–98.
- [97] S. Klainerman and A. Majda, *Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids*, Comm. Pure Appl. Math. **34** (4) (1981), 481–524.
- [98] S. Klainerman and A. Majda, *Compressible and incompressible fluids*, Comm. Pure Appl. Math. **35** (5) (1982), 629–651.
- [99] H.-O. Kreiss, *Problems with different time scales for partial differential equations*, Comm. Pure Appl. Math. **33** (3) (1980), 399–439.
- [100] M. Lachowicz, *On the initial layer and the existence theorem for the nonlinear Boltzmann equation*, Math. Methods Appl. Sci. **9** (3) (1987), 342–366.
- [101] O.E. Lanford, III, *Time evolution of large classical systems*, Dynamical Systems, Theory and Applications (Recontres, Battelle Res. Inst., Seattle, WA, 1974), Lecture Notes in Phys., Vol. 38, Springer, Berlin (1975), 1–111.

- [102] J. Leray, *Etude de diverses équations intégrales nonlinéaires et de quelques problèmes que pose l'hydrodynamique*, J. Math. Pures Appl. **12** (1933), 1–82.
- [103] J. Leray, *Essai sur les mouvements plans d'un liquide visqueux emplissant l'espace*, Acta. Math. **63** (1934), 193–248.
- [104] J. Leray, *Essai sur les mouvements plans d'un liquide visqueux qui limitent des parois*, J. Math. Pures Appl. **13** (1934), 331–418.
- [105] D. Levermore and N. Masmoudi, *From the Boltzmann equation to an incompressible Navier–Stokes–Fourier system*, Preprint (2005).
- [106] C.K. Lin, *On the incompressible limit of the compressible Navier–Stokes equations*, Comm. Partial Differential Equations **20** (3/4) (1995), 677–707.
- [107] J.-L. Lions, *Some Methods in the Mathematical Analysis of Systems and Their Control*, Kexue Chubanshe (Science Press), Beijing (1981).
- [108] P.-L. Lions, *Mathematical Topics in Fluid Mechanics, Vol. 1. Incompressible Models*, The Clarendon Press Oxford University Press, New York (1996).
- [109] P.-L. Lions, *Mathematical Topics in Fluid Mechanics, Vol. 2, Compressible Models*, The Clarendon Press Oxford University Press, New York (1998). Oxford Science Publications.
- [110] P.-L. Lions, *Bornes sur la densité pour les équations de Navier–Stokes compressibles isentropiques avec conditions aux limites de Dirichlet*, C. R. Acad. Sci. Paris Sér. I Math. **328** (8) (1999), 659–662.
- [111] P.-L. Lions and N. Masmoudi, *Incompressible limit for a viscous compressible fluid*, J. Math. Pures Appl. (9) **77** (6) (1998), 585–627.
- [112] P.-L. Lions and N. Masmoudi, *On a free boundary barotropic model*, Ann. Inst. H. Poincaré Anal. Non Linéaire **16** (3) (1999), 373–410.
- [113] P.-L. Lions and N. Masmoudi, *Une approche locale de la limite incompressible*, C. R. Acad. Sci. Paris Sér. I Math. **329** (5) (1999), 387–392.
- [114] P.-L. Lions and N. Masmoudi, *From the Boltzmann equations to the equations of incompressible fluid mechanics. I*, Arch. Ration. Mech. Anal. **158** (3) (2001), 173–193.
- [115] P.-L. Lions and N. Masmoudi, *From the Boltzmann equations to the equations of incompressible fluid mechanics. II*, Arch. Ration. Mech. Anal. **158** (3) (2001), 195–211.
- [116] P.-L. Lions and N. Masmoudi, *Homogenization of the Euler system in a 2D porous medium*, J. Math. Pures Appl. (9) **84** (1) (2005), 1–20.
- [117] A. Majda, *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*, Applied Mathematical Sciences, Vol. 53, Springer-Verlag, New York (1984).
- [118] A. Majda, *Introduction to PDEs and Waves for the Atmosphere and Ocean*, Courant Lecture Notes in Mathematics, Vol. 9, New York University Courant Institute of Mathematical Sciences, New York (2003).
- [119] A.J. Majda and A.L. Bertozzi, *Vorticity and Incompressible Flow*, Cambridge Texts in Applied Mathematics, Vol. 27, Cambridge University Press, Cambridge (2002).
- [120] C. Marchioro, *On the inviscid limit for a fluid with a concentrated vorticity*, Comm. Math. Phys. **196** (1) (1998), 53–65.
- [121] C. Marchioro and M. Pulvirenti, *Vortex Methods in Two-Dimensional Fluid Dynamics*, Lecture Notes in Physics, Vol. 203, Springer-Verlag, Berlin (1984).
- [122] C. Marchioro and M. Pulvirenti, *Mathematical Theory of Incompressible Nonviscous Fluids*, Applied Mathematical Sciences, Vol. 96, Springer-Verlag, New York (1994).
- [123] N. Masmoudi, *The Euler limit of the Navier–Stokes equations, and rotating fluids with boundary*, Arch. Rational Mech. Anal. **142** (4) (1998), 375–394.
- [124] N. Masmoudi, *Asymptotic problems and compressible–incompressible limit*, Advances in Mathematical Fluid Mechanics (Paseky, 1999), Springer-Verlag, Berlin (2000), 119–158.
- [125] N. Masmoudi, *Ekman layers of rotating fluids: The case of general initial data*, Comm. Pure Appl. Math. **53** (4) (2000), 432–483.
- [126] N. Masmoudi, *From Vlasov–Poisson system to the incompressible Euler system*, Comm. Partial Differential Equations **26** (9/10) (2001), 1913–1928.
- [127] N. Masmoudi, *Incompressible, inviscid limit of the compressible Navier–Stokes system*, Ann. Inst. H. Poincaré Anal. Non Linéaire **18** (2) (2001), 199–224.
- [128] N. Masmoudi, *Some uniform elliptic estimates in a porous medium*, C. R. Math. Acad. Sci. Paris (12) **339** (2004), 849–854.

- [129] N. Masmoudi, *Two remarks about the inviscid limit of the Navier–Stokes system*, Preprint (2006).
- [130] N. Masmoudi and K. Nakanishi, *Nonrelativistic limit from Maxwell–Klein–Gordon and Maxwell–Dirac to Poisson–Schrödinger*, Int. Math. Res. Not. **13** (2003), 697–734.
- [131] N. Masmoudi and K. Nakanishi, *From the Klein–Gordon–Zakharov system to the nonlinear Schrödinger equation*, J. Hyperbolic Differ. Equ. **2** (4) (2005), 975–1008.
- [132] N. Masmoudi and L. Saint-Raymond, *From the Boltzmann equation to the Stokes–Fourier system in a bounded domain*, Comm. Pure Appl. Math. **56** (9) (2003), 1263–1293.
- [133] G. Métivier and S. Schochet, *The incompressible limit of the non-isentropic Euler equations*, Arch. Ration. Mech. Anal. **158** (1) (2001), 61–90.
- [134] G. Métivier and S. Schochet, *Averaging theorems for conservative systems and the weakly compressible Euler equations*, J. Differential Equations **187** (1) (2003), 106–183.
- [135] A. Mikelic, *Homogenization of nonstationary Navier–Stokes equations in a domain with a grained boundary*, Ann. Mat. Pura Appl. (4) **158** (1991), 167–179.
- [136] S. Mischler, *On the initial boundary value problem for the Vlasov–Poisson–Boltzmann system*, Comm. Math. Phys. **210** (2) (2000), 447–466.
- [137] G. Nguetseng, *A general convergence result for a functional related to the theory of homogenization*, SIAM J. Math. Anal. **20** (3) (1989), 608–623.
- [138] G. Nguetseng, *Asymptotic analysis for a stiff variational problem arising in mechanics*, SIAM J. Math. Anal. **21** (6) (1990), 1394–1414.
- [139] A. Novotný and I. Straskraba, *Introduction to the Mathematical Theory of Compressible Flow*, Oxford Lecture Series in Mathematics and Its Applications, Vol. 27, Oxford University Press, Oxford (2004).
- [140] O.A. Oleinik, *On the system of Prandtl equations in boundary-layer theory*, Dokl. Akad. Nauk SSSR **150** (1963), 28–31.
- [141] O.A. Oleinik, *On the mathematical theory of boundary layer for an unsteady flow of incompressible fluid*, J. Appl. Math. Mech. **30** (1967) (1966), 951–974.
- [142] O.A. Oleinik and V.N. Samokhin, *Mathematical models in boundary layer theory*, Applied Mathematics and Mathematical Computation, Vol. 15, Chapman & Hall/CRC, Boca Raton, FL (1999).
- [143] M. Paicu, *Étude asymptotique pour les fluides anisotropes en rotation rapide dans le cas périodique*, J. Math. Pures Appl. (9) **83** (2) (2004), 163–242.
- [144] J. Pedlovsky, *Geophysical Fluid Dynamics*, Springer-Verlag, New York (1979).
- [145] L. Prandtl, *Boundary layer*, Verhandlung Internationalen Mathematiker-Kongresses, Heidelberg (1904), 484–491.
- [146] G. Raugel and G.R. Sell, *Navier–Stokes equations in thin 3D domains. III. Existence of a global attractor*, Turbulence in Fluid Flows, IMA Vol. Math. Appl., Vol. 55, Springer-Verlag, New York (1993), 137–163.
- [147] F. Rousset, *Large mixed Ekman–Hartmann boundary layers in magnetohydrodynamics*, Nonlinearity **17** (2) (2004), 503–518.
- [148] F. Rousset, *Stability of large Ekman boundary layers in rotating fluids*, Arch. Ration. Mech. Anal. **172** (2) (2004), 213–245.
- [149] L. Saint-Raymond, *Convergence of solutions to the Boltzmann equation in the incompressible Euler limit*, Arch. Ration. Mech. Anal. **166** (1) (2003), 47–80.
- [150] M. Sammartino and R.E. Caflisch, *Zero viscosity limit for analytic solutions, of the Navier–Stokes equation on a half-space. I. Existence for Euler and Prandtl equations*, Comm. Math. Phys. **192** (2) (1998), 433–461.
- [151] M. Sammartino and R.E. Caflisch, *Zero viscosity limit for analytic solutions of the Navier–Stokes equation on a half-space. II. Construction of the Navier–Stokes solution*, Comm. Math. Phys. **192** (2) (1998), 463–491.
- [152] E. Sánchez-Palencia, *Nonhomogeneous Media and Vibration Theory*, Springer-Verlag, Berlin (1980).
- [153] S. Schochet, *The compressible Euler equations in a bounded domain: Existence of solutions and the incompressible limit*, Comm. Math. Phys. **104** (1) (1986), 49–75.
- [154] S. Schochet, *Fast singular limits of hyperbolic PDEs*, J. Differential Equations **114** (2) (1994), 476–512.
- [155] S. Schochet, *The mathematical theory of low Mach number flows*, M2AN Math. Model. Numer. Anal. **39** (3) (2005), 441–458.
- [156] Y. Sone, *Kinetic Theory and Fluid Dynamics*, Modeling and Simulation in Science, Engineering and Technology, Birkhäuser, Boston, MA (2002).

- [157] H. Spohn, *Boltzmann hierarchy and Boltzmann equation*, Kinetic Theories and the Boltzmann Equation (Montecatini, 1981), Springer-Verlag, Berlin (1984), 207–220.
- [158] R.S. Strichartz, *Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations*, Duke Math. J. **44** (3) (1977), 705–714.
- [159] H.S.G. Swann, *The convergence with vanishing viscosity of nonstationary Navier–Stokes flow to ideal flow in R^3* , Trans. Amer. Math. Soc. **157** (1971), 373–397.
- [160] L. Tartar, *Incompressible fluid flow in a porous medium: Convergence of the homogenization process*, Nonhomogeneous Media and Vibration Theory, Sánchez-Palencia, E., ed. (1980), 368–377.
- [161] L. Tartar, *H-measures, a new approach for studying homogenisation, oscillations and concentration effects in partial differential equations*, Proc. Roy. Soc. Edinburgh Sect. A **115** (3/4) (1990), 193–230.
- [162] R. Temam, *On the Euler equations of incompressible perfect fluids*, J. Functional Analysis **20** (1) (1975), 32–43.
- [163] R. Temam, *Navier–Stokes Equations and Nonlinear Functional Analysis*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1983).
- [164] R. Temam and X. Wang, *The convergence of the solutions of the Navier–Stokes equations to that of the Euler equations*, Appl. Math. Lett. **10** (5) (1997), 29–33.
- [165] R. Temam and X. Wang, *Boundary layers associated with incompressible Navier–Stokes equations: The noncharacteristic boundary case*, J. Differential Equations **179** (2) (2002), 647–686.
- [166] S. Ukai, *The incompressible limit and the initial layer of the compressible Euler equation*, J. Math. Kyoto Univ. **26** (2) (1986), 323–331.
- [167] S. Ukai and K. Asano, *The Euler limit and initial layer of the nonlinear Boltzmann equation*, Hokkaido Math. J. **12** (3, part 1) (1983), 311–332.
- [168] C. Villani, *A review of mathematical problems in collisional kinetic theory*, Handbook of Mathematical Fluid Dynamics, Vol. II, Friedlander, S. and Serre, D., eds., North-Holland, Amsterdam (2003).
- [169] X. Wang, *A Kato type theorem on zero viscosity limit of Navier–Stokes flows*, Indiana Univ. Math. J. **50** (Special Issue) (2001), 223–241. Dedicated to Professors Ciprian Foias and Roger Temam (Bloomington, IN, 2000).
- [170] X. Wang, *Infinite Prandtl number limit of Rayleigh–Bénard convection*, Comm. Pure Appl. Math. **57** (10) (2004), 1265–1282.
- [171] W. Wolibner, *Un théorème sur l’existence du mouvement plan d’un fluide parfait, homogène, incompressible, pendant un temps infiniment long*, Math. Z. **37** (1) (1933), 698–726.
- [172] S. Wu, *Well-posedness in Sobolev spaces of the full water wave problem in 2-D*, Invent. Math. **130** (1997), 39–72.
- [173] Z. Xin and L. Zhang, *On the global existence of solutions to the Prandtl’s system*, Adv. Math. **181** (1) (2004), 88–133.
- [174] H.-T. Yau, *Relative entropy and hydrodynamics of Ginzburg–Landau models*, Lett. Math. Phys. **22** (1) (1991), 63–80.
- [175] V.I. Yudovich, *Non-stationary flows of an ideal incompressible fluid*, Z. Vychisl. Mat. i Mat. Fiz. **3** (1963), 1032–1066.
- [176] R.K. Zeytounian, *Asymptotic Modelling of Fluid Flow Phenomena*, Fluid Mechanics and Its Applications, Vol. 64, Kluwer, Dordrecht (2002).
- [177] R.K. Zeytounian, *Theory and Applications of Viscous Fluid Flows*, Springer-Verlag, Berlin (2004).

CHAPTER 4

Notes on Hyperbolic Systems of Conservation Laws and Transport Equations

Camillo De Lellis

*Institut für Mathematik, Universität Zürich,
Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
E-mail: delellis@math.unizh.ch*

Contents

1. Introduction	279
1.1. The Keyfitz and Kranzer system	279
1.2. Bressan's compactness conjecture	280
1.3. Ambrosio's renormalization theorem	281
1.4. Well-posedness for the Keyfitz and Kranzer system	282
1.5. Renormalization conjecture for nearly incompressible BV fields	283
1.6. Plan of the paper	284
2. Preliminaries	284
2.1. Notation	284
2.2. Measure theory	285
2.3. Approximate continuity and approximate jumps	287
2.4. BV functions	288
2.5. Caccioppoli sets and coarea formula	289
2.6. The Volpert chain rule	289
2.7. Alberti's rank-one theorem	290
3. DiPerna-Lions theory for nearly incompressible flows	290
3.1. Lagrangian flows	291
3.2. Nearly incompressible fields and fields with the renormalization property	294
3.3. Existence and uniqueness of solutions to transport equations	298
3.4. Stability of solutions to transport equations	302
3.5. Existence, uniqueness and stability of regular Lagrangian flows	304
4. Commutator estimates and Ambrosio's renormalization theorem	308
4.1. Difference quotients of BV functions	310
4.2. Commutator estimate	312
4.3. Bouchut's lemma and Alberti's lemma	317
4.4. Proof of Theorem 4.1	320

HANDBOOK OF DIFFERENTIAL EQUATIONS

Evolutionary Equations, volume 3

Edited by C.M. Dafermos and E. Feireisl

© 2007 Elsevier B.V. All rights reserved

5. Existence, uniqueness, and stability for the KeyPtz and Kranzer system	321
5.1. Proof of Theorem 5.5	324
5.2. Renormalized entropy solutions are entropy solutions	326
5.3. Proof of Proposition 5.9	330
6. Blow-up of the BV norm for the KeyPtz and Kranzer system	334
6.1. Preliminary lemmas	335
6.2. Proof of Theorem 6.1	338
6.3. Proof of Proposition 6.5	344
7. Partial regularity and trace properties of solutions to transport equations	349
7.1. Anzellotti's weak trace for measure $\mathbb{D}$ divergence bounded vector fields	350
7.2. Further properties of Anzellotti's weak trace	354
7.3. Change of variables for traces	357
7.4. Proof of Theorem 7.1	359
7.5. Proof of Theorem 7.4	363
8. Bressan's compactness conjecture and the renormalization conjecture for nearly incompressible BV vector fields	364
8.1. Absolutely continuous and jump parts of the measure $D \cdot (\rho h(w)B)$	366
8.2. Proof of Proposition 8.4 and concentration of commutators	368
8.3. Proof of Theorem 8.10	369
9. Tangential sets of BV vector fields	374
Acknowledgements	380
References	380

1. Introduction

The aim of this chapter is to give an account of some recent results about transport equations with variable BV coefficients, and their applications to a class of hyperbolic systems of conservation laws in several space dimensions. Besides collecting results which are scattered in the literature, it has been my intention to give a self-contained and more readable reference, and to provide details, remarks and connections barely mentioned in the original papers.

1.1. The Keyfitz and Kranzer system

We start by considering the following system of equations:

$$\begin{cases} \partial_t u^i + \sum_{\alpha=1}^m \partial_{x_\alpha} (g^\alpha(|u|) u^i) = 0, \\ u^i(0, \cdot) = \bar{u}^i(\cdot), \end{cases} \quad (1)$$

where $u = (u^1, \dots, u^k) : \mathbb{R}_t^+ \times \mathbb{R}_x^m \rightarrow \mathbb{R}^k$ is the unknown vector map, $\bar{u} = (\bar{u}^1, \dots, \bar{u}^k)$ the initial data, and $g^\alpha : \mathbb{R} \rightarrow \mathbb{R}$ are given (sufficiently smooth) scalar functions. In one space dimension (1) was first studied by Keyfitz and Kranzer in [34] and later on by several other authors, as a prototypical example of a nonstrict hyperbolic system; see for instance [28, 31, 35, 38]. Indeed, in the one-dimensional terminology, the hyperbolicity of (1) degenerates at the origin (see for instance [23], Section 7.2).

However, the Keyfitz and Kranzer system has many features. In particular, it can be formally reduced to a scalar conservation law and a system of transport equations with variable coefficients. More precisely, if u is smooth and solves (1), then $\rho := |u|$ solves

$$\begin{cases} \partial_t \rho + D_x \cdot (\rho g(\rho)) = 0, \\ \rho(0, \cdot) = |\bar{u}|(\cdot), \end{cases} \quad (2)$$

and, if in addition $|u| > 0$, then $\theta := u/|u|$ solves

$$\begin{cases} \partial_t \theta + g(\rho) \cdot D_x \theta = 0, \\ \theta(0, \cdot) = \frac{\bar{u}}{|\bar{u}|(\cdot)}. \end{cases} \quad (3)$$

One can use this observation to produce solutions to (1). However, as it is well known, even starting from extremely regular initial data, solutions of (2) develop singularities in finite time, and one cannot hope to get better than BV regularity. Thus, in order to construct solutions in the way described above, one has to face the problem of solving transport equations

$$\begin{cases} \partial_t \theta(t, x) + b(t, x) \cdot D_x \theta(t, x) = 0, \\ \theta(0, x) = \bar{\theta}(x), \end{cases} \quad (4)$$

when b is quite irregular.

From now on, we will say that a distributional solution u of (1) is a *renormalized entropy solution* if $\rho := |u|$ solves, in the sense of Kruzkov, the scalar law (2) (see Definitions 5.1 and 5.4).

1.2. Bressan's compactness conjecture

In [17] Bressan showed that in two space dimensions renormalized entropy solutions might lead to an ill-posed Cauchy problem for bounded initial data. However he conjectured that this does not happen when the absolute value of the initial data are in BV_{loc} . In particular, in order to show the existence of renormalized entropy solutions to (1) when $|\bar{u}| \in L^\infty \cap BV$ and $|\bar{u}|^{-1} \in L^\infty$, he advanced the following conjecture (see also [18,19]).

CONJECTURE 1.1 (Bressan's compactness conjecture). Let $b_n: \mathbb{R}_t \times \mathbb{R}_x^m \rightarrow \mathbb{R}^m$, $n \in \mathbb{N}$, be smooth maps and denote by Φ_n the solutions of the ODEs

$$\begin{cases} \frac{d}{dt} \Phi_n(t, x) = b_n(t, \Phi_n(t, x)), \\ \Phi_n(0, x) = x. \end{cases} \quad (5)$$

Assume that $\|b_n\|_\infty + \|\nabla b_n\|_{L^1}$ is uniformly bounded and that the fluxes Φ_n are nearly incompressible, i.e., that

$$C^{-1} \leq \det(\nabla_x \Phi_n(t, x)) \leq C \quad \text{for some constant } C > 0. \quad (6)$$

Then the sequence $\{\Phi_n\}$ is strongly precompact in L^1_{loc} .

An affirmative answer to this conjecture leads immediately to the existence of renormalized entropy solutions of (1) when $C \geq |\bar{u}| \geq c > 0$ and $\bar{u} \in BV$. Indeed, assume that these assumptions hold and consider the Kruzkov solution ρ of (2). It is well known that $\rho \in BV_{\text{loc}}$ and $C \geq \rho \geq c > 0$. Thus, $g(\rho)$ is also $BV_{\text{loc}} \cap L^\infty$. It is not difficult to see that we can approximate $b := g(\rho)$ and ρ with two sequences $\{b_n\}$ and $\{\rho_n\}$ of smooth functions such that

- (i) $\|b_n\|_{BV} + \|b_n\|_\infty$ is uniformly bounded;
- (ii) $C_1 \geq \rho_n \geq c_1 > 0$ for some constant c_1 ;
- (iii) $\partial_t \rho_n + D_x \cdot (b_n \rho_n) = 0$.

If we set $\bar{\theta} := \bar{u}/\bar{\rho}$, then we can solve

$$\begin{cases} \partial_t \theta_n(t, x) + b_n(t, x) \cdot D_x \theta_n(t, x) = 0, \\ \theta(0, x) = \bar{\theta}(x), \end{cases} \quad (7)$$

with the classical method of characteristics. If we let Φ_n be as in (5), then the continuity equations of (iii), condition (ii) and the standard maximum principle for transport equations with smooth coefficients imply the existence of a constant C such that (6) holds. At this stage we could use Conjecture 1.1 to show that θ_n converges locally strongly to a function θ (up to subsequences). This strong convergence implies that $u := \theta \rho$ is a renormalized entropy solution.

1.3. Ambrosio's renormalization theorem

In the recent ground-breaking paper [2] (see also [5]), Ambrosio has shown well-posedness of

$$\begin{cases} \partial_t \theta(t, x) + b(t, x) \cdot D_x \theta(t, x) = 0, \\ \theta(0, x) = \bar{\theta}(x), \end{cases} \quad (8)$$

under the assumptions that $b \in BV$ and $D_x \cdot b$ is a bounded function.

The result of Ambrosio uses the theory of renormalized solutions, first introduced by DiPerna and Lions in [27] (in that paper the authors proved, among other results, the well-posedness of (8) under the assumptions $b \in L^\infty \cap W^{1,1}$ and $D_x \cdot b \in L^\infty$).

The core of Ambrosio's well-posedness theorem is a new Renormalization lemma. In order to understand its content, consider first a smooth vector field B in $\Omega \subset \mathbb{R}^d$ and a smooth scalar function u such that $B \cdot Du = 0$. For any smooth function β the classical chain rule yields

$$B \cdot D(\beta(u)) = B \cdot [\beta'(u) Du] = 0.$$

Next assume that $B \in BV$, that the divergence $D \cdot B$ is an absolutely continuous measure, and that $u \in L^\infty$. Then the expression

$$D \cdot (uB) - u D \cdot B$$

makes sense distributionally, and can be taken as a definition of $B \cdot Du$. Ambrosio's renormalization theorem states that the conclusion

$$0 = B \cdot D(\beta(u)) := D \cdot (\beta(u)B) - \beta(u) D \cdot B \quad \forall \beta \in C^1(\mathbb{R})$$

holds even under these much weaker assumptions.

Assume now that $b \in BV$, $D_x \cdot b \in L^1$ and u is a bounded weak solution of the transport equation $\partial_t u + b \cdot Du = 0$ with initial data $\bar{u}$. More precisely, assume that

$$\begin{aligned} & \int_0^\infty \int_{\mathbb{R}^m} u(t, x) \{ \partial_t \varphi(t, x) + b(t, x) \cdot D \varphi(t, x) \\ & \quad - [D_x \cdot b](t, x) \varphi(t, x) \} dt dx \\ & = - \int_{\mathbb{R}^m} \bar{u}(x) \varphi(0, x) dx \end{aligned}$$

for every smooth compactly supported test function φ . Applying Ambrosio's renormalization theorem to the field $B = (1, b): \mathbb{R}^+ \times \mathbb{R}^m \rightarrow \mathbb{R} \times \mathbb{R}^m$, we infer that $\beta(u)$ solves the corresponding Cauchy problem with initial data $\beta(\bar{u})$ (actually a technical step is needed in order to conclude that $\beta(u)$ has initial data $\beta(\bar{u})$; see Sections 3 and 4). If in addition we

have the bounds $b \in L^\infty$ and $D_x \cdot b \in L^\infty$, the equation satisfied by $\beta(\bar{u})$ can be used (for special choices of β) to derive estimates and comparison principles, via standard Gronwall-type arguments. These comparison principles are indeed enough to show uniqueness and stability for weak solutions of (8).

A byproduct of the renormalization property is that solutions of (8) are stable even under approximation of the coefficient b . In the DiPerna-Lions theory this is used to conclude existence, stability, and compactness properties for the ODEs with coefficients b . Therefore Ambrosio's result can be used to infer that Bressan's compactness conjecture holds when we replace the bound (6) with the stronger assumption

$$-C \leq D_x \cdot B \leq C. \quad (9)$$

1.4. Well-posedness for the Keyfitz and Kranzer system

Though presently there is no general proof of Bressan's compactness conjecture, it is still possible to use Ambrosio's renormalization theorem to show existence of renormalized entropy solutions when $|\bar{u}| \in BV_{\text{loc}}$. The difference with respect to Bressan's compactness conjecture is that in this specific case one can take advantage of an additional information. Indeed, if ρ is a Kruzhkov solution of the scalar law (2), then the coefficient $b := g(\rho)$ has a solution of the continuity equation which, besides being bounded from above and from below, also enjoys BV regularity. This information is missing in the assumptions of Conjecture 1.1.

Basically Ambrosio's renormalization lemma is powerful enough to provide a DiPerna-Lions theory for transport equations with $BV \cap L^\infty$ coefficients which possess a BV non-negative solution ρ of the continuity equation. As shown in [4], this yields well-posedness for the Keyfitz and Kranzer system when $|\bar{u}| \in BV_{\text{loc}} \cap L^\infty$ (in particular it also allows to drop the unnatural assumption $|\bar{u}| \geq c > 0$). More precisely, for every $\bar{u}$ with $|\bar{u}| \in BV_{\text{loc}} \cap L^\infty$ there exists a unique renormalized entropy solution of (1). Moreover, if a sequence of initial data $\bar{u}_n$ converges to $\bar{u}$ and $\|\bar{u}_n\|_\infty + \|\bar{u}_n\|_{BV_{\text{loc}}}$ is uniformly bounded, then the corresponding renormalized entropy solutions converge.

This result raises the following natural question: Is system (1) well posed in BV ? In other words, when the whole initial data $\bar{u}$ (and not only its absolute value $|\bar{u}|$) is in BV , does the renormalized entropy solution enjoy BV regularity? The answer to this question is no to a large extent. More precisely, in [25] it has been shown that, in three space dimensions, for every g which is not constant there exist bounded renormalized entropy solutions of (1) which are not in BV_{loc} but have BV initial data. These examples can be produced by starting from initial data which are arbitrarily close (both in L^∞ and BV norm) to a constant different from 0. Thus, the lack of BV regularity nor is a 'large data' effect, neither is due to the degeneracy of the hyperbolicity of the system at the origin. In two space dimensions similar examples can be produced for a large class of fluxes g .

The same 'irregularity' also holds for general entropy solutions. Indeed in [25] it is shown that, when the convex hull of the essential image of $\bar{u}$ does not contain the origin, any bounded admissible solution of (1) with BV regularity necessarily coincides with the renormalized entropy solution.

1.5. Renormalization conjecture for nearly incompressible BV fields

Though we can prove the well-posedness of (1) bypassing Conjecture 1.1, this conjecture remains a challenging and interesting open problem in the theory of transport equations with nonsmooth coefficients. Presently we are able to show it only under some technical assumptions (the most general result concerning Bressan's compactness conjecture is contained in [10]). One interesting case in which we are able to show Conjecture 1.1 is when we assume that the singular part of the measure $D_x \cdot b$ is concentrated on a set of codimension 1.

Our approach to Conjecture 1.1 is again through the theory of renormalized solutions à la DiPerna and Lions. Indeed, though we drop the assumption $D_x \cdot b \in L^1$, it is possible to use nonnegative solutions of the continuity equation $\partial_t \rho + D_x \cdot (\rho b) = 0$ to build a theory of renormalized solutions. In this framework, in [4] we proposed a renormalization lemma for nearly incompressible BV coefficients which is a natural generalization of Ambrosio's renormalization theorem. More precisely, we have the following conjecture.

CONJECTURE 1.2 (Renormalization conjecture). Let $\Omega \subset \mathbb{R}^d$ be an open set. Assume $B \in BV \cap L^\infty(\Omega, \mathbb{R}^d)$ and $\rho \in L^\infty(\Omega)$ satisfy $D \cdot (\rho B) = 0$ and $\rho \geq C > 0$. Then, for every $u \in L^\infty(\Omega)$ such that $D \cdot (\rho u B) = 0$ and for every $\beta \in C^1$, we have $D \cdot (\rho \beta(u) B) = 0$.

This conjectured chain rule leads naturally to investigate coupling between bounded functions and measures. Recently, in [6] the authors have shown trace theorems and regularity properties for ρ and u , coming from the equations $D \cdot (\rho B) = 0$ and $D \cdot (\rho u B) = 0$. In particular, it turns out that ρ and u possess a suitably strong notion of trace on hypersurfaces which are transversal to B . In [10] we combine these trace properties with Ambrosio's renormalization theorem to show Conjecture 8.2 when the singular part of the measure $D \cdot B$ is concentrated on a set of codimension 1.

In the general case, we decompose the measure $D \cdot B$ into the part which is absolutely continuous with respect to the Lebesgue measure and the singular part, denoted respectively by $D^a \cdot B$ and $D^s \cdot B$. Further, we follow [24] and decompose $D^s \cdot B$ into a jump part $D^j \cdot B$, concentrated on a set of codimension 1, and a Cantor part $D^c \cdot B$ (see Section 2 and [11] for the details). It turns out that $D^j \cdot B$ is concentrated on the set where the BV field B has jump-singularities (the jump set J_B), whereas the measure $D^c \cdot B$ is a singular measure of fractal type which is less singular than $D^j \cdot B$: More precisely, $|D^c \cdot B|(\Sigma) = 0$ for every set Σ of codimension 1 with finite Hausdorff measure. In this framework, the result mentioned in the previous paragraph can be restated as

- Conjecture 1.2 has a positive answer when $D^c \cdot B = 0$.

However, the results of [6] and [10] allow to handle a more general case. Indeed, one can define a notion of transversality between the measure $D^c \cdot B$ and the field B . In [6] the authors showed that, when $D^c \cdot B$ and B are transversal, ρ and u are approximately continuous $|D^c \cdot B|$ -almost everywhere. In [10] we prove a new renormalization result, showing that Conjecture 1.2 holds whenever ρ and u are approximately continuous $|D^c \cdot B|$ -a.e. Thus we conclude that Conjecture 1.2 holds whenever $D^c \cdot B$ and B are transversal.

Unfortunately it is possible to show BV fields for which $D^c \cdot B$ and B are *not* transversal (see Section 9 and [10]). However it is not clear whether this can happen under the additional hypothesis that B is nearly incompressible.

1.6. Plan of the paper

In Section 2 we collect facts about measure theory and BV functions which will be relevant to our purposes, together with appropriate references on where to find their proofs. In Section 3 we develop the DiPerna-Lions theory for nearly incompressible fields. In Section 4 we prove Ambrosio's renormalization theorem and in Section 5 we use this theorem and the DiPerna-Lions theory to address the existence, uniqueness and stability of renormalized entropy solutions to the Keyfitz and Kranzer system. In Section 6 we show that the BV norm of renormalized entropy solutions blow up in a large number of cases.

In the last three sections we address the most recent results on the renormalization conjecture. Section 7 contains the trace properties and partial regularity of solutions to transport equations proved in [6]. Section 8 follows [10] and shows Conjecture 1.2 under the assumption that ρ and u are approximately continuous $|D^c \cdot B|$ -a.e. Finally, Section 9 contains an example of [10]: A planar BV vector field for which $D^c \cdot B$ and B are not transversal.

2. Preliminaries

In this section we will collect some preliminary facts about measure theory and BV functions. Most of them can be found in the monograph [11].

2.1. Notation

When $\Omega \subset \mathbb{R}^d$, we will denote by id the identity map $\text{id}: \Omega \ni x \rightarrow x \in \mathbb{R}^d$. If $x_1, \dots, x_d$ is a standard system of coordinates on $\mathbb{R}^d$ we denote by $\{e_i\}_{i=1, \dots, d}$ the standard unit orthonormal vector fields such that $x = \sum_i x_i e_i$. If A and B are $k \times n$ and $n \times m$ matrices, $A \cdot B$ will denote the usual product ($k \times m$) matrix, whereas A^T will denote the transpose of the matrix A . Vectors will usually be considered as $n \times 1$ matrices and therefore, if a and b are vectors, $a^T \cdot b$ is the usual scalar product. With a slight abuse of notation we will simply write $a \cdot b$, and similarly, if a and b are vectors and A is a matrix, we will use $a \cdot A \cdot b$ in place of $a^T \cdot A \cdot b$.

Given a vector valued map $B: \Omega \rightarrow \mathbb{R}^k$ and some system of coordinates on $\mathbb{R}^k$, with $\{e_i\}_{i=1, \dots, k}$ orthonormal vectors, we will denote by B^i the scalar function given by $e_i \cdot B$. Whereas the subscript B_j will be always used to denote the element of a sequence $\{B_j\}_{j \in \mathbb{N}}$ of maps.

If $E \subset \mathbb{R}^d$ then we denote by $\mathbb{1}_E$ the function given by

$$\mathbb{1}_E(x) := \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Given $x \in \mathbb{R}^d$ and $r > 0$ we denote by $B_r(x)$ the ball of $\mathbb{R}^d$ centered at x of radius r . $\mathcal{L}^d$ denotes the Lebesgue d -dimensional measure, $\mathcal{H}^k$ denotes the usual Hausdorff k -dimensional measure, and we set $\omega_d := \mathcal{L}^d(B_1(0))$.

When μ is a measure and A a μ -measurable set, we denote by $\mu \llcorner A$ the measure given by

$$\mu \llcorner A(B) = \mu(A \cap B).$$

In many case, we will deal with the Lebesgue measure $\mathcal{L}^d$ restricted on some measurable set $\Omega \subset \mathbb{R}^d$. When it will be clear from the context, to simplify the notation we will use $\mathcal{L}^d$ in place of $\mathcal{L}^d \llcorner \Omega$.

If μ on A is a measure and $f: A \rightarrow B$ is a measurable function, then we denote by $f_{\#}\mu$ the usual push-forward of μ , that is, the measure on B defined by

$$\int \varphi \, d[f_{\#}\mu] = \int \varphi(f(x)) \, d\mu(x) \quad \text{for every } \varphi \in C_c(A).$$

When μ is Radon (vector-valued) measure, $|\mu|$ denotes its total variation measure. Moreover, if $E \subset \Omega$ is a Borel set and μ a Radon measure on Ω such that $|\mu|(\Omega \setminus E) = 0$, then we say that μ is concentrated on E .

We say that $\eta \in C_c^\infty(\mathbb{R}^d)$ is a standard kernel if $\int \eta = 1$. Moreover, for any $\varepsilon > 0$ we denote by η_ε the function defined by $\eta_\varepsilon(x) := \varepsilon^{-d} \eta(x/\varepsilon)$. If Ω is an open subset of $\mathbb{R}^d$ and $f \in L^1(\Omega)$, then we denote by $f * \eta_\varepsilon$ the function $(f \mathbb{1}_\Omega) * \eta_\varepsilon$.

If $T \in \mathcal{D}'(\Omega)$, then we denote by $\langle T, \varphi \rangle$ the value of T on the test function $\varphi \in C_c^\infty(\Omega)$. Moreover, if η is as above, we set

$$T * \eta_\delta(y) := \langle T, \eta_\delta(\cdot - y) \rangle$$

for every $y \in \Omega$ such that $\eta_\delta(\cdot - y)$ is compactly supported in Ω . In particular, if $\tilde{\Omega} \subset\subset \Omega$ and δ is sufficiently small, $T * \eta_\delta$ defines a distribution in $\mathcal{D}'(\tilde{\Omega})$.

2.2. Measure theory

We now recall the following elementary results in measure theory (see for instance Proposition 1.62(b) of [11]).

PROPOSITION 2.1. *Let $\{\mu_n\}_n$ be a sequence of Radon measures on $\Omega \subset \mathbb{R}^d$, which converge weakly* to μ and assume that $|\mu_n|$ converge weakly* to λ . Then $\lambda \geq |\mu|$. Moreover, if E is a compact set or a bounded open set such that $\lambda(\partial E) = 0$, then $\mu_n(E) \rightarrow \mu(E)$.*

PROPOSITION 2.2. *Let μ be a Radon measure on Ω , $\eta \in C_c^\infty(\mathbb{R}^d)$ be a standard kernel supported in the unit ball, and $\{\eta_\delta\}_\delta$ the corresponding standard family of mollifiers. Then, for any $\tilde{\Omega} \subset\subset \Omega$, $\mu * \eta_\delta$ converges weakly* to μ in $\tilde{\Omega}$ and $|\mu * \eta_\delta|$ converges weakly* to $|\mu|$ in $\tilde{\Omega}$.*

Let μ be a Radon $\mathbb{R}^k$ -valued measure on Ω . By the Lebesgue decomposition theorem, μ has a unique decomposition into *absolutely continuous part* μ^a and *singular part* μ^s with respect to Lebesgue measure $\mathcal{L}^d$. Further, by the Radon–Nikodym theorem there exists a unique $f \in L^1_{\text{loc}}(\Omega, \mathbb{R}^k)$ such that $\mu^a = f \mathcal{L}^d$.

One can further decompose μ^s as follows.

PROPOSITION 2.3 (Decomposition of the singular part). *If $|\mu^s|$ vanishes on any $\mathcal{H}^{d-1}$ -negligible set, then μ^s can be uniquely written as a sum $\mu^c + \mu^j$ of two measures such that*

- (a) $\mu^c(A) = 0$ for every Borel set A with $\mathcal{H}^{d-1}(A) < +\infty$;
- (b) $\mu^j = f \mathcal{H}^{d-1} \llcorner J_\mu$ for some Borel set J_μ σ -finite with respect to $\mathcal{H}^{d-1}$.

The proof of this proposition is analogous to the proof of decomposition of derivatives of BV functions (and indeed in this case the decompositions coincide), see Proposition 3.92 of [11]. In this proof, the Borel set J_μ is defined as

$$J_\mu := \left\{ x \in \Omega \mid \limsup_{r \downarrow 0} \frac{|\mu|(B_r(x))}{r^{d-1}} > 0 \right\}. \quad (10)$$

These measures will be called, respectively, *jump part* and *Cantor part* of the measure μ . Sometimes we will use the notation μ^d for the measure $\mu^a + \mu^c$ (here the superscript diffused stays for diffused).

For $B \in L^1_{\text{loc}}(\Omega, \mathbb{R}^k)$ we denote by $DB = (D_i B^j)_{ij}$ the derivative in the sense of distributions of B , i.e. the $\mathbb{R}^{k \times d}$ -valued distribution defined by

$$\langle D_i B^j, \varphi \rangle := - \int_{\Omega} B^j \frac{\partial \varphi}{\partial x_i} dx \quad \forall \varphi \in C_c^\infty(\Omega), \quad 1 \leq i \leq d, 1 \leq j \leq k.$$

When $\Omega \subset \mathbb{R}^d$ and $k = d$, we denote by $D \cdot B$ the distribution $\sum_i D_i B^i$. We have the following lemma.

LEMMA 2.4. *Let $\Omega \subset \mathbb{R}^d$ and let $B \in L^\infty(\Omega, \mathbb{R}^d)$ be such that $D \cdot B$ is a Radon measure. Then $D \cdot B \ll \mathcal{H}^{d-1}$.*

Thanks to this lemma, for any bounded vector field B such that $D \cdot B$ is a Radon measure, we can apply the decomposition of Definition 2.3 to $D \cdot B$. Therefore we will denote by $D^a \cdot B$, $D^c \cdot B$ and $D^j \cdot B$ respectively the absolutely continuous part, Cantor part and jump part of $D \cdot B$. Moreover we will sometimes use $D^s \cdot B$ for $D^c \cdot B + D^j \cdot B$ and $D^d \cdot B$ for $D^a \cdot B + D^c \cdot B$.

PROOF OF LEMMA 2.4. We will show that $|(D \cdot B)(B_r(x))| \leq \|B\|_\infty \omega_{d-1} r^{d-1}$ for every ball $B_r(x) \subset \subset \Omega$. This implies the claim by a standard covering argument (see for instance Theorem 2.56 of [11]). Therefore let $x \in \Omega$ be given and fix a smooth nonnegative kernel $\eta \in C_c^\infty(\mathbb{R}^d)$. Consider $\mu_\varepsilon := D \cdot (B * \eta_\varepsilon) = (D \cdot B) * \eta_\varepsilon$. Then $\mu_\varepsilon \rightharpoonup^* D \cdot B$ on any set

$\tilde{\Omega} \subset \subset \Omega$. Note that for any fixed $B_r(x) \subset \subset \Omega$ we have

$$\begin{aligned} |\mu_\varepsilon(B_r(x))| &= \left| \int_{B_r(x)} D_x \cdot (B * \eta_\varepsilon)(x) dx \right| \\ &= \left| \int_{\partial B_r(x)} B * \eta_\varepsilon \cdot \nu \right| \leq \|B * \eta_\varepsilon\|_{\infty} \omega_{d-1} r^{d-1} \leq \|B\|_{\infty} \omega_{d-1} r^{d-1}. \end{aligned}$$

Define $S \subset]0, \text{dist}(x, \partial\Omega)[$ as the set of radii ρ such that $|D \cdot B|(\partial B_r(x)) > 0$, which is at most countable. Since $\mu_\varepsilon \rightharpoonup^* D \cdot B$, for any $r \in]0, \text{dist}(x, \partial\Omega)[\setminus S$ we have

$$| [D \cdot B](B_r(x)) | = \lim_{\varepsilon \downarrow 0} |\mu_\varepsilon(B_r(x))| \leq \|B\|_{\infty} \omega_{d-1} r^{d-1}.$$

Moreover, since S is at most countable, for any $r \in S$ there exists $\{r_n\} \subset]0, \text{dist}(x, \partial\Omega)[\setminus S$ such that $r_n \uparrow r$. Therefore

$$| [D \cdot B](B_r(x)) | = \lim_{r_n \uparrow r} | [D \cdot B](B_{r_n}(x)) | \leq \|B\|_{\infty} \omega_{d-1} r^{d-1}. \quad \square$$

2.3. Approximate continuity and approximate jumps

The L^1 -approximate discontinuity set $S_B \subset \Omega$ of a locally summable $B : \Omega \rightarrow \mathbb{R}^k$ and the Lebesgue limit are defined as follows: $x \notin S_B$ if and only if there exists $z \in \mathbb{R}^k$ satisfying

$$\lim_{r \downarrow 0} r^{-d} \int_{B_r(x)} |B(y) - z| dy = 0.$$

The vector z , if it exists, is unique and denoted by $\tilde{B}(x)$, the Lebesgue limit of B at x . It is easy to check that the set S_B is Borel and that $\tilde{B}$ is a Borel function in its domain (see Section 3.6 of [11] for details). By Lebesgue differentiation theorem the set S_B is Lebesgue negligible and $\tilde{B} = B$ $\mathcal{L}^d$ -a.e. in $\Omega \setminus S_B$.

In a similar way one can define the L^1 -approximate jump set $J_B \subset S_B$, by requiring the existence of $a, b \in \mathbb{R}^k$ with $a \neq b$ and of a unit vector ν such that

$$\lim_{r \downarrow 0} r^{-d} \int_{B_r^+(x, \nu)} |B(y) - a| dy = 0, \quad \lim_{r \downarrow 0} r^{-d} \int_{B_r^-(x, \nu)} |B(y) - b| dy = 0,$$

where

$$\begin{aligned} B_r^+(x, \nu) &:= \{y \in B_r(x) : \langle y - x, \nu \rangle > 0\}, \\ B_r^-(x, \nu) &:= \{y \in B_r(x) : \langle y - x, \nu \rangle < 0\}. \end{aligned} \quad (11)$$

The triplet (a, b, ν) , if it exists, is unique up to a permutation of a and b and a change of sign of ν , and denoted by $(B^+(x), B^-(x), \nu(x))$, where $B^\pm(x)$ are called *Lebesgue one-sided limits* of B at x . It is easy to check that the set J_B is Borel and that $B^\pm$ and ν can be chosen to be Borel functions in their domain (see again Section 3.6 of [11] for details).

2.4. BV functions

DEFINITION 2.5 (BV functions). We say that $B \in L^1(\Omega; \mathbb{R}^k)$ has bounded variation in Ω , and we write $B \in BV(\Omega; \mathbb{R}^k)$, if DB is representable by an $\mathbb{R}^{k \times d}$ -valued measure, still denoted by DB , with finite total variation in Ω .

It is a well known fact that for $B \in BV$ one has $D_i B^j \ll \mathcal{H}^{d-1}$ (for instance it follows directly from Lemma 2.4 applied to the vector field $U = B^j e_i$). Therefore we can apply the decomposition of Section 2.1 to the measure DB and we will use the notation $D^a B$, $D^c B$ and $D^j B$, respectively for the absolutely continuous part, Cantor part and jump part of DB . Moreover we will denote by $D^s B$ and $D^d B$ respectively the measures $D^c B + D^j B$ and $D^a B + D^c B$.

Next we recall the basic properties of $\mathbb{R}^k$ -valued BV functions defined in an open set $\Omega \subset \mathbb{R}^d$.

First of all we need the definition of rectifiable sets.

DEFINITION 2.6 (Countably $\mathcal{H}^{d-1}$ -rectifiable sets). We say that $\Sigma \subset \mathbb{R}^d$ is countably $\mathcal{H}^{d-1}$ -rectifiable if there exist (at most) countably many C^1 embedded hypersurfaces $\Gamma_i \subset \mathbb{R}^d$ such that

$$\mathcal{H}^{d-1}\left(\Sigma \setminus \bigcup_i \Gamma_i\right) = 0.$$

A Borel map $\nu: \Sigma \rightarrow \mathbb{S}^{d-1}$ is normal to Σ if $\nu(x)$ is normal to Γ_i for $\mathcal{H}^{d-1}$ -a.e. $x \in \Gamma_i \cap \Sigma$.

Denoting by $\zeta \otimes \xi$ the linear map from $\mathbb{R}^d$ to $\mathbb{R}^k$ defined by $v \mapsto \zeta \langle \xi, v \rangle$, the following structure theorem holds (see for instance Theorem 3.77 and Proposition 3.92 of [11]).

THEOREM 2.7 (BV structure theorem). If $B \in BV_{\text{loc}}(\Omega, \mathbb{R}^k)$, then $\mathcal{H}^{d-1}(S_B \setminus J_B) = 0$ and J_B is a countably $\mathcal{H}^{d-1}$ -rectifiable set. Moreover,

$$D^j B = (B^+ - B^-) \otimes \nu \mathcal{H}^{d-1} \llcorner J_B, \quad (12)$$

and ν is normal to Σ .

As a corollary, since $D^a B$ and $D^c B$ are both concentrated on $\Omega \setminus S_B$, we conclude that $|D^a B| + |D^c B| = |D^d B|$ -a.e. x is a Lebesgue point for B , with value $\tilde{B}(x)$. The space of functions of special bounded variation (denoted by SBV) is defined as follows.

DEFINITION 2.8 (SBV). Let $\Omega \subset \mathbb{R}^d$ be an open set. The space $SBV(\Omega, \mathbb{R}^m)$ is the set of all $u \in BV(\Omega, \mathbb{R}^m)$ such that $D^c u = 0$.

2.5. Caccioppoli sets and coarea formula

We say that $A \subset \Omega$ is a Caccioppoli set if $\mathbb{1}_A \in BV(\Omega)$. Then, as a particular case of Theorem 2.7, we conclude that there exists a rectifiable set F such that:

- for every $x \notin F$ the Lebesgue limit of $\mathbb{1}_A$ is either 0 or 1;
- $\mathcal{H}^{d-1}$ -a.e. $x \in F$ is an approximate jump point for $\mathbb{1}_A$ such that $\mathbb{1}_A^+(x) = 1$, $\mathbb{1}_A^-(x) = 0$ and ν is normal to F ;
- $D^j \mathbb{1}_A = \nu \mathcal{H}^{d-1} \llcorner F$.

F is called the essential boundary of A and denoted by $\partial^* A$ (see Section 3.5 of [11]). ν is called the approximate exterior unit normal to A . An additional important fact is that $D^c \mathbb{1}_A = D^a \mathbb{1}_A = 0$. More precisely, we have (cf. with Theorem 3.59 of [11]) the theorem.

THEOREM 2.9 (De Giorgi's rectifiability theorem). *If A is a Caccioppoli set, then $D\mathbb{1}_A = D^j \mathbb{1}_A = \nu \mathcal{H}^{d-1} \llcorner \partial^* A$.*

Thus, $\mathcal{H}^{d-1}(A) = |D\mathbb{1}_A|(\Omega) < \infty$.

A second important tool of the theory of BV functions is the coarea formula. Before stating it, we introduce the following notation. Assume that $[a, b] \ni t \mapsto \mu_t$ is a map which takes values on the space of $\mathbb{R}^k$ -valued measures. We say that this map is weakly* measurable if for every test function $\varphi \in C_c(\Omega, \mathbb{R}^k)$, the map $t \mapsto \int \varphi d\mu_t$ is measurable. If $\int |\mu_t|(\Omega) dt$ is finite, then we denote by $\int \mu_t dt$ the measure μ defined by

$$\int \varphi d\mu := \int \left(\int \varphi d\mu_t \right) dt.$$

Then we have (cf. with Theorem 3.40 of [11]).

THEOREM 2.10 (Coarea formula). *Let $u \in BV(\Omega)$ be a scalar BV function. For $t \geq 0$ we set $\Omega_t := \{u > t\}$ and for $t < 0$ we set $\Omega_t := \{u < t\}$. Then Ω_t is a Caccioppoli set for $\mathcal{L}^1$ -a.e. t , $t \mapsto D\mathbb{1}_{\Omega_t}$ is a weakly* measurable, and $\int |D\mathbb{1}_{\Omega_t}|(\Omega) dt < \infty$. Moreover,*

$$Du = \int_0^\infty D\mathbb{1}_{\Omega_t} - \int_0^\infty D\mathbb{1}_{\Omega_{-t}}, \quad (13)$$

$$|Du| = \int_{-\infty}^\infty \mathcal{H}^{d-1} \llcorner \partial^* \Omega_t dt. \quad (14)$$

2.6. The Vol'pert chain rule

Next, note that if $B \in BV(\Omega, \mathbb{R}^k)$ and $H \in W^{1,\infty}(\mathbb{R}^k, \mathbb{R}^m)$ then $H \circ B \in BV_{\text{loc}}(\Omega, \mathbb{R}^m)$. Indeed, let $\{B_n\}_n$ be any sequence of smooth functions such that $B_n \rightarrow B$ strongly in L^1 and

$$\limsup_{n \uparrow \infty} \int_\Omega |\nabla B_n(x)| dx < \infty.$$

Clearly, $H \circ B_n \rightarrow H \circ B$ strongly in L^1 and

$$\begin{aligned} \limsup_{n \uparrow \infty} \int_{\Omega} |\nabla[H \cdot B_n](x)| \, dx &= \limsup_{n \uparrow \infty} \int_{\Omega} |\nabla H(B_n(x)) \cdot \nabla B(x)| \, dx \\ &\leq \|\nabla H\|_{\infty} \limsup_{n \uparrow \infty} \int_{\Omega} |\nabla B_n(x)| \, dx < \infty. \end{aligned}$$

Therefore $D[H \cdot B]$ is a Radon measure. In addition, if $H \in C^1$, then the following chain rule, first proved by Volpert, holds (see Theorem 3.96 of [11]).

THEOREM 2.11. *Let $u \in BV(\Omega, \mathbb{R}^k)$ and $H \in C^1(\mathbb{R}^k, \mathbb{R}^m)$. Then*

$$D[H \circ u] = [\nabla H \circ \tilde{u}] \cdot D^d u + \{[H(u^+) - H(u^-)] \otimes \nu\} \mathcal{H}^{d-1} \llcorner J_u. \quad (15)$$

REMARK 2.12. In [7] the authors proved a suitable extension of Theorem 2.11 to $H \in W^{1,\infty}$. In what follows we will sometimes consider the measure $D[H \circ u]$ for H which indeed are $W^{1,\infty}$ but not C^1 . However we will not need the general result of [7], since in all the cases considered in this paper we will be able to use some *à hoc* considerations.

2.7. Alberti's rank-one theorem

In [1] Alberti proved the following deep result.

THEOREM 2.13 (Alberti's rank-one theorem). *Let $B \in BV_{\text{loc}}(\Omega, \mathbb{R}^k)$. Then there exist Borel functions $\xi: \Omega \rightarrow \mathbb{S}^{d-1}$, $\zeta: \Omega \rightarrow \mathbb{S}^{k-1}$ such that*

$$D^s B = \zeta \otimes \xi |D^s B|. \quad (16)$$

Clearly, if we replace $D^s B$ with $D^j B$ in (16), this conclusion can be easily drawn from Theorem 2.7. However, in order to prove the same for the *full* singular part of DB , many new interesting ideas were introduced in [1] (see also [26] for a recent description of Alberti's proof).

3. DiPerna–Lions theory for nearly incompressible flows

In this section we develop a theory *à la* DiPerna and Lions for transport equations and ordinary differential equations, in which the usual assumption of boundedness of the divergence of the coefficients is replaced by a control on the Jacobian (or by the existence of a solution of the continuity equation which is bounded away from 0 and ∞).

3.1. Lagrangian flows

DEFINITION 3.1. Let $b \in L^\infty([0, \infty[\times \mathbb{R}^m, \mathbb{R}^m)$. A map $\Phi : [0, \infty[\times \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a regular Lagrangian flow for b if

- (a) for $\mathcal{L}^1$ -a.e. t we have $|\{x : \Phi(t, x) \in A\}| = 0$ for every Borel set A with $|A| = 0$;
- (b) the following identity is valid in the sense of distributions

$$\begin{cases} \partial_t \Phi(t, x) = b(t, \Phi(t, x)), \\ \Phi(0, x) = x. \end{cases} \quad (17)$$

The identity (17) in the sense of distributions means that for every $\psi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^m, \mathbb{R}^m)$ we have

$$\begin{aligned} & \int_{\mathbb{R}^m} \psi(0, x) \cdot x \, dx + \int_0^\infty \int_{\mathbb{R}^m} \Phi(t, x) \cdot \partial_t \psi(t, x) \, dt \, dx \\ &= - \int_0^\infty \int_{\mathbb{R}^m} \psi(t, x) \cdot b(t, \Phi(t, x)) \, dt \, dx. \end{aligned} \quad (18)$$

Note that assumption (a) guarantees that $b(t, \Phi(t, x))$ is well defined. More precisely, if $\hat{b}(t, x) = b(t, x)$ for $\mathcal{L}^{m+1}$ -a.e. (t, x) , then $\hat{b}(t, \Phi(t, x)) = b(t, \Phi(t, x))$ for $\mathcal{L}^{m+1}$ -a.e. (t, x) .

Moreover, it is easy to check that if Φ is a regular Lagrangian flow and $\Psi(t, x) = \Phi(t, x)$ for $\mathcal{L}^{m+1}$ -a.e. (t, x) , then Ψ is as well a regular Lagrangian flow.

The following lemma has a standard proof.

LEMMA 3.2. Let Φ be a regular Lagrangian flow. Then, $\Phi(\cdot, x) \in W_{\text{loc}}^{1,\infty}([0, \infty[)$ for $\mathcal{L}^m$ -a.e. x and, if we denote by Φ_x the Lipschitz function such that $\Phi_x(t) = \Phi(t, x)$ for $\mathcal{L}^1$ -a.e. t , then:

- $\text{Lip}(\Phi_x) \leq \|b\|_\infty$.
- $\Phi_x(0) = x$.
- $\Phi'_x(t) = b(t, \Phi_x(t))$ for $\mathcal{L}^1$ -a.e. t .

The following is an easy corollary of Lemma 3.2.

COROLLARY 3.3. Let Φ be a regular Lagrangian flow. Then, for any Borel set A and $\mathcal{L}^1$ -a.e. $T > 0$ we have

$$\int_A |\Phi(T, x) - x| \, dx \leq \|b\|_\infty T |A|. \quad (19)$$

From now on we denote by μ_Φ the measure $(\text{id}, \Phi)_\# \mathcal{L}^{m+1} \llcorner ([0, \infty[\times \mathbb{R}^m)$, that is the push forward via the map $(t, x) \mapsto (t, \Phi(t, x))$ of the Lebesgue $(m+1)$ -dimensional measure on $[0, \infty[\times \mathbb{R}^m$. Thus

$$\int_{[0, \infty[\times \mathbb{R}^m} \psi(t, x) \, d\mu_\Phi(t, x) = \int_{[0, \infty[\times \mathbb{R}^m} \psi(t, \Phi(t, x)) \, d\mathcal{L}^{m+1}(t, x)$$

for every $\psi \in C^c(\mathbb{R} \times \mathbb{R}^m)$.

Having introduced μ_Φ , (a) is equivalent to

$$\mu_\Phi \ll \mathcal{L}^{m+1}. \quad (20)$$

Thus for every regular Lagrangian flow Φ there exists a $\rho \in L^1_{\text{loc}}([0, \infty[\times \mathbb{R}^n)$ such that $\mu_\Phi = \rho \mathcal{L}^{m+1}$.

DEFINITION 3.4. This ρ will be called the *density* of the flow Φ , and by definition it satisfies the following change of variables identity

$$\int \psi(t, \Phi(t, x)) dt dx = \int \psi(t, x) \rho(t, x) dt dx \quad (21)$$

for every test function $\psi \in L^\infty$ and with bounded support.

The next proposition shows the connections between regular Lagrangian flows and solutions of transport and continuity equations with coefficient b .

PROPOSITION 3.5. *Let Φ be a regular Lagrangian flow for a field b .*

- (i) *Let $\bar{\zeta} \in L^\infty(\mathbb{R}^n)$ and consider the measure μ on $[0, \infty[\times T$ given by $(\text{id}, \Phi)_\#(\bar{\zeta} \mathcal{L}^{m+1})$, that is,*

$$\int \varphi(t, x) d\mu(t, x) = \int_A \varphi(t, \Phi(t, x)) \bar{\zeta}(x) dt dx \quad \text{for every Borel set } A.$$

Then there exists $\zeta \in L^1_{\text{loc}}([0, \infty[\times \mathbb{R}^m)$ such that $\mu = \zeta \mathcal{L}^{m+1}$. Moreover, ζ satisfies the following equation in the sense of distributions:

$$\begin{cases} \partial_t \zeta + D_x \cdot (\zeta b) = 0, \\ \zeta(0, \cdot) = \bar{\zeta}. \end{cases} \quad (22)$$

- (ii) *Let ρ be the density of the flow Φ . If $u \in L^\infty([0, T[\times \mathbb{R}^m)$ and $\bar{u} \in L^\infty(\mathbb{R}^m)$ satisfy the identity*

$$u(t, \Phi(t, x)) = \bar{u}(x) \quad \text{for } \mathcal{L}^{m+1}\text{-a.e. } (t, x), \quad (23)$$

then the following equation holds in the sense of distributions

$$\begin{cases} \partial_t(\rho u) + D_x \cdot (\rho u b) = 0, \\ u(0, \cdot) \rho(0, \cdot) = \bar{u}. \end{cases} \quad (24)$$

Thus, as a particular case of this proposition, we get the usual continuity equation satisfied by the density ρ of flows of regular vector fields:

$$\begin{cases} \partial_t \rho + D_x \cdot (\rho b) = 0, \\ \rho(0, \cdot) = 1. \end{cases} \quad (25)$$

PROOF OF PROPOSITION 3.5. First of all note that (ii) follows from (i). Indeed, let u and $\bar{u}$ be given as in (ii). Set $\bar{\zeta} := \bar{u}$ and $\zeta := u\rho$. For every L^∞ function with bounded support φ we have

$$\begin{aligned} \int u(t, x) \rho(t, x) \varphi(t, x) \, dt \, dx &= \int u(t, \Phi(t, x)) \varphi(t, \Phi(t, x)) \, dt \, dx \\ &= \int \bar{u}(x) \varphi(t, \Phi(t, x)) \, dt \, dx. \end{aligned}$$

Thus, if μ is defined as in (i), then $\zeta \mathcal{L}^{m+1} = \mu$. Therefore (i) gives (22), from which we get (24).

We now come to the proof of (i). First of all note that

$$\begin{aligned} |\mu(A)| &= \left| \int \bar{\zeta}(x) \mathbb{1}_A(t, \Phi(t, x)) \, dt \, dx \right| \\ &\leq \|\bar{\zeta}\|_\infty \int \mathbb{1}_A(t, \Phi(t, x)) \, dt \, dx \leq \|\bar{\zeta}\|_\infty \int_A \rho(t, x) \, dt \, dx. \end{aligned}$$

Since $\rho \in L^1_{\text{loc}}$, this means that μ is absolutely continuous. Therefore there exists an L^1_{loc} function ζ such that $\mu = \zeta \mathcal{L}^{m+1}$. Now, let $\psi \in C^\infty(\mathbb{R} \times \mathbb{R}^m)$ be any given test function. Our goal is to show that

$$\begin{aligned} & - \int_{[0, \infty[\times \mathbb{R}^n} \zeta(t, x) (\partial_t \psi(t, x) + b(t, x) \cdot \nabla_x \psi(t, x)) \, dx \, dt \\ &= \int_{\mathbb{R}^n} \bar{\zeta}(x) \psi(0, x) \, dx. \end{aligned} \tag{26}$$

By definition, the left-hand side of (26) is equal to

$$- \int_{\mathbb{R}^n} \bar{\zeta}(x) \left[\int_0^\infty (\partial_t \psi(t, \Phi(t, x)) + \nabla_x \psi(t, \Phi(t, x)) \cdot b(t, \Phi(t, x))) \, dt \right] dx. \tag{27}$$

We conclude the proof by showing that, for any x for which the conclusion of Lemma 3.2 applies, we have

$$-\psi(0, x) = \int_0^\infty (\partial_t \psi(t, \Phi_x(t)) + \nabla_x \psi(t, \Phi_x(t)) \cdot \Phi'_x(t)) \, dt.$$

For such x the integral in t in (27) is given by

$$\int_0^\infty (\partial_t \psi(t, \Phi_x(t)) + \nabla_x \psi(t, \Phi_x(t)) \cdot \Phi'_x(t)) \, dt.$$

Since Φ_x is Lipschitz and ψ is a smooth function, $\psi(\cdot, \Phi_x(\cdot))$ is a Lipschitz function of t . Therefore, $\psi(\cdot, \Phi_x(\cdot))$ and $\Phi_x(\cdot)$ are both differentiable at $\mathcal{L}^1$ -a.e. t , and the identity given by the usual chain rule

$$\partial_t \psi(t, \Phi_x(t)) + \nabla_x \psi(t, \Phi_x(t)) \cdot \Phi'_x(t) = \frac{d}{dt} (\psi(t, \Phi_x(t)))$$

is valid for a.e. t . Moreover, note that

- $\psi(0, \Phi_x(0)) = \psi(0, x)$;
- $\psi(T, \Phi_x(T)) = 0$ for T large enough, since η has bounded support.

Therefore we conclude

$$\int_0^\infty (\partial_t \psi(t, \Phi(t, x)) + \nabla_x \psi(t, \Phi(t, x)) \cdot b(t, \Phi(t, x))) dt = -\psi(0, x). \quad (28)$$

□

3.2. Nearly incompressible fields and fields with the renormalization property

DEFINITION 3.6. We say that a field $b \in L^\infty([0, \infty[\times \mathbb{R}^m, \mathbb{R}^m)$ is nearly incompressible if there exists a function $\rho \in L^\infty([0, \infty[\times \mathbb{R}^m)$ and a positive constant C such that $C^{-1} \leq \rho \leq C$ and

$$\partial_t \rho + D_x \cdot (\rho b) = 0 \quad (29)$$

in the sense of distributions.

The following lemma has a standard proof.

LEMMA 3.7. If ρ is bounded and satisfies (29), then, after possibly modifying it on a set of measure zero, $[0, 1] \ni t \mapsto \rho(t, \cdot) \in L^\infty$ is a weakly* continuous map.

REMARK 3.8. As a consequence of Lemma 3.7 we get the following useful fact. Given any $\zeta \in C_c^\infty([0, \infty[)$ with $\int \zeta = 1$, if we denote by $\{\zeta_\varepsilon\}$ the standard family of mollifiers generated by ζ , then the functions

$$\int_0^\infty \zeta_\varepsilon(t) \rho(t, x) dt$$

converge weakly* in L^∞ to $\rho(0, \cdot)$.

PROOF OF LEMMA 3.7. We claim that

(Cl) For every $\varphi \in C_c^\infty(\mathbb{R}^m)$ the functions

$$f_\varphi^T(t) := \int_{\mathbb{R}^m} \frac{1}{T} \int_t^{T+t} \rho(s, x) \varphi(x) ds dx$$

are uniformly continuous.

This claim implies the lemma. Indeed, let $\varphi \in C_c^\infty(\mathbb{R}^m)$. Then from (Cl) we conclude that $\{f_\varphi^T\}_{0 < T < 1}$ is precompact in $C([0, R])$ for every $R > 0$. Let f denote any limit of a subsequence $\{f_\varphi^{T_k}\}$ with $T_k \downarrow 0$. Then we have

$$\int f(t)\psi(t) dt = \int \rho(t, x)\varphi(x)\psi(t) dt dx$$

for every $\psi \in C_c^\infty(\mathbb{R})$. Therefore we conclude that f_φ^T is converging (uniformly on compact sets) to a unique $f_\varphi^0 \in C([0, \infty])$, as $T \rightarrow 0$.

It is clear that $|f_\varphi^0(t)| \leq \|\rho\|_\infty \|\varphi\|_{L^1}$ and that $f_{a\varphi+b\psi}^0(t) = af_\varphi^0(t) + bf_\psi^0(t)$. Therefore for each t there exists a unique $\rho_t \in L^\infty$ such that

$$\int \rho_t(x)\varphi(x) dx = f_\varphi^0(t) \quad \text{for every } \varphi \in C_c^\infty(\mathbb{R}^m).$$

Since $C_c^\infty(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$, the map $t \mapsto \rho_t$ is weakly* continuous. Moreover, for any test function $\psi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^m)$ we have

$$\int \rho_t(x)\psi(t, x) dt dx = \int \rho(t, x)\psi(t, x) dt dx.$$

It remains to show (Cl). Therefore, let $\varphi \in C_c^\infty(\mathbb{R}^n)$ be any given test function. For every $0 < T < 1$ consider

$$\chi_T(t) := \begin{cases} \frac{t}{T} & \text{for } t \in [0, T], \\ 1 & \text{for } t \in [T, 1], \\ 2-t & \text{for } t \in [1, 2], \\ 0 & \text{for } t \geq 2. \end{cases}$$

Set $\psi_T(t, \tau, x) := \chi_T(\tau - t)\varphi(x)$. It is not difficult to see that

$$\int \rho(\tau, x)(\partial_t \psi_T(t, \tau, x) + b(\tau, x) \cdot \nabla_x \psi_T(t, \tau, x)) d\tau dx = 0,$$

from which we get

$$\begin{aligned} f_T(t) &= \frac{1}{T} \int_t^{t+T} \int_{\mathbb{R}^m} \rho(\tau, x)\varphi(x) dx d\tau \\ &= \int_{t+1}^{t+2} \int_{\mathbb{R}^n} \rho(\tau, x)\varphi(x) dx d\tau \\ &\quad - \int_0^\infty \int_{\mathbb{R}^m} \rho(\tau, x)\chi_T(\tau - t)\nabla\varphi(x) \cdot b(\tau, x) dx d\tau. \end{aligned}$$

From this identity we easily conclude that $\{f_T\}_{0 < T < 1}$ is uniformly continuous. $\square$

DEFINITION 3.9. We say that a pair $b \in L^\infty([0, \infty[\times \mathbb{R}^m, \mathbb{R}^m)$, $\rho \in L^\infty([0, \infty[\times \mathbb{R}^m)$ have the renormalization property if ρ satisfies (29) and the following property holds:

(R) For every $T > 0$ and for every bounded u which solves

$$\begin{cases} \partial_t(\rho u) + D_x \cdot (u \rho b) = 0, \\ [u\rho](0, \cdot) = \rho(0, \cdot) \bar{u}, \\ [u\rho](T, \cdot) = \rho(T, \cdot) \hat{u}. \end{cases} \quad (30)$$

$v := u^2$ solves

$$\begin{cases} \partial_t(\rho v) + D_x \cdot (v \rho b) = 0, \\ [v\rho](0, \cdot) = \rho(0, \cdot) \bar{u}^2, \\ [v\rho](T, \cdot) = \rho(T, \cdot) \hat{u}^2. \end{cases} \quad (31)$$

In the previous definition $\rho(0, \cdot)$ and $\rho(T, \cdot)$ are the traces of ρ given by Lemma 3.7, and the identity (30) means that for every test function $\varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}^m)$ we have

$$\begin{aligned} & \int_{[0, \infty[\times \mathbb{R}^m} \rho(t, x) u(t, x) (\partial_t \varphi(t, x) + b(t, x) \cdot \nabla \varphi(t, x)) \, dt \, dx \\ &= \int_{\mathbb{R}^m} (\rho(T, x) \hat{u}(x) \varphi(T, x) - \rho(0, x) \bar{u}(x) \varphi(0, x)) \, dx. \end{aligned}$$

The following proposition holds.

PROPOSITION 3.10. Assume that (b, ρ) have the renormalization property. Then:

(GR) For every finite family of bounded solutions $\{u^i\}_{i=1, \dots, N}$ of

$$\begin{cases} \partial_t(\rho u^i) + D_x \cdot (u^i \rho b) = 0, \\ [u^i \rho](0, \cdot) = \rho(0, \cdot) \bar{u}^i, \\ [u^i \rho](T, \cdot) = \rho(T, \cdot) \hat{u}^i, \end{cases} \quad (32)$$

and any $H \in C(\mathbb{R}^N)$, $v := H(u)$ solves

$$\begin{cases} \partial_t(\rho v) + D_x \cdot (v \rho b) = 0, \\ [v\rho](0, \cdot) = \rho(0, \cdot) H(\bar{u}), \\ [v\rho](T, \cdot) = \rho(T, \cdot) H(\hat{u}). \end{cases} \quad (33)$$

PROOF. Note that the claim is always true when H is a linear function. Moreover, since $u^1 u^2 = ((u^1 + u^2)^2 - (u^1)^2 - (u^2)^2)/2$, from the renormalization property (R) we conclude that

$$(GR) \text{ holds for } N = 2 \text{ and } H(u^1, u^2) = u^1 u^2. \quad (34)$$

Using inductively (34) we get that

$$(GR) \text{ holds whenever } H \text{ is a polynomial.} \quad (35)$$

In order to prove the general case, let u and H be given as in the statement of the proposition. By Stone-Weierstrass there exists a sequence of polynomials $H_k : \mathbb{R}^N \rightarrow \mathbb{R}$ such that $H_k \rightarrow H$ uniformly on $\overline{B}_{\|u\|_\infty}(0) \subset \mathbb{R}^N$. From (35) we get

$$\begin{cases} \partial_t(\rho H_k(u)) + D_x \cdot (H_k(u) \rho b) = 0, \\ [H_k(u) \rho](0, \cdot) = \rho(0, \cdot) H_k(\bar{u}), \\ [H_k(u) \rho](T, \cdot) = \rho(T, \cdot) H_k(\hat{u}), \end{cases} \quad (36)$$

and letting $k \uparrow \infty$ we conclude (33). $\square$

COROLLARY 3.11. *Let b a bounded nearly incompressible vector field with the renormalization property, and assume that ρ is as in Definitions 3.6 and 3.9. If ζ is any other function such that $0 < C^{-1} \leq \zeta \leq C$ and $\partial_t \zeta + D_x \cdot (\zeta b) = 0$, then (GR) also holds with ζ in place of ρ .*

This corollary justifies the following definition.

DEFINITION 3.12. We say that a bounded nearly incompressible vector field b has the *renormalization property* if there exists a ρ as in Definition 3.6 such that the pair (b, ρ) has the renormalization property of Definition 3.9.

PROOF OF COROLLARY 3.11. Let $\{u^i\}_{i=1, \dots, N}$ be any given solutions of

$$\begin{cases} \partial_t(\zeta u^i) + D_x \cdot (u^i \zeta b) = 0, \\ [u^i \zeta](0, \cdot) = \zeta(0, \cdot) \bar{u}^i, \\ [u^i \zeta](T, \cdot) = \zeta(T, \cdot) \hat{u}^i. \end{cases} \quad (37)$$

Next, let $v^{n+1} := \zeta / \rho$, $\bar{v}^{n+1} := \zeta(0, \cdot) / \rho(0, \cdot)$, and $\hat{v}^{n+1} := \zeta(T, \cdot) / \rho(T, \cdot)$. Then define $v^i := u^i / v^{n+1}$, $\bar{v}^i := \bar{u}^i / \bar{v}^{n+1}$ and $\hat{v}^i := \hat{u}^i / \hat{v}^{n+1}$. Note that

$$\begin{cases} \partial_t(\rho v^i) + D_x \cdot (v^i \rho b) = 0, \\ [v^i \rho](0, \cdot) = \rho(0, \cdot) \bar{v}^i, \\ [v^i \rho](T, \cdot) = \rho(T, \cdot) \hat{v}^i. \end{cases} \quad (38)$$

Given $H \in C(\mathbb{R}^N)$, we define $\widehat{H} \in C(\mathbb{R}^{N+1})$ by $\widehat{H}(v) := v^{n+1} H(v^1 v^{n+1}, \dots, v^n v^{n+1})$. Since (GR) holds, we conclude

$$\begin{cases} \partial_t(\rho \widehat{H}(v)) + D_x \cdot (\widehat{H}(v) \rho b) = 0, \\ [\widehat{H}(v) \rho](0, \cdot) = \rho(0, \cdot) \widehat{H}(\bar{v}), \\ [\widehat{H}(v) \rho](T, \cdot) = \rho(T, \cdot) \widehat{H}(\hat{v}). \end{cases} \quad (39)$$

On the other hand, from the definitions of v and $\widehat{H}$, we have

$$\begin{aligned} \rho \widehat{H}(v) &= \zeta H(u), & \rho(0, \cdot) \widehat{H}(\bar{v}) &= \zeta(0, \cdot) H(\bar{u}) \quad \text{and} \\ \rho(T, \cdot) \widehat{H}(\hat{v}) &= \zeta(T, \cdot) H(\hat{u}). \end{aligned} \quad \square$$

3.3. Existence and uniqueness of solutions to transport equations

PROPOSITION 3.13. *Assume b is a bounded vector field and ρ is a nonnegative function which satisfies (29). Then for every bounded $\bar{u}$ there exists a solution of*

$$\begin{cases} \partial_t(\rho u) + D_x \cdot (u \rho b) = 0, \\ [u \rho](0, \cdot) = \bar{u} \bar{\rho}. \end{cases} \quad (40)$$

Assume, moreover, that the pair (b, ρ) has the renormalization property. If u_1 and u_2 solve

$$\begin{cases} \partial_t(\rho u_i) + D_x \cdot (u_i \rho b) = 0, \\ [u_i \rho](0, \cdot) = \bar{u}_i \rho(0, \cdot), \end{cases} \quad (41)$$

and $\bar{u}_1 \geq \bar{u}_2$, then $\rho u_1 \geq \rho u_2$.

The following are easy corollaries of Proposition 3.13.

COROLLARY 3.14. *If b is a bounded nearly incompressible vector field with the renormalization property and ρ is as in Definition 3.6, then for every bounded $\bar{u}$ there exists a unique bounded solution u of (40). Moreover, after possibly changing u on a set of measure zero, the map $t \mapsto u(t, \cdot)$ is continuous in the strong topology of L^1_{loc} .*

COROLLARY 3.15. *Let $\bar{\zeta} \in L^\infty(\mathbb{R}^m)$. If b is a bounded nearly incompressible vector field with the renormalization property, then there exists a unique bounded distributional solution ζ of*

$$\begin{cases} \partial_t \zeta + D_x \cdot (\zeta b) = 0, \\ \zeta(0, \cdot) = \bar{\zeta}. \end{cases} \quad (42)$$

Moreover, if $\bar{\zeta}$ is bounded away from zero, so is ζ .

This justifies the following definition.

DEFINITION 3.16. Let b be a bounded nearly incompressible vector field with the renormalization property. Then the *density* generated by b is the unique solution of

$$\begin{cases} \partial_t \rho + D_x \cdot (\rho b) = 0, \\ \rho(0, \cdot) = 1. \end{cases} \quad (43)$$

Moreover note that, if Φ is a regular Lagrangian flow for b , then the density of Φ coincides with the density generated by b .

The proof of the comparison principle of Proposition 3.13 is an easy consequence of the following lemma.

LEMMA 3.17. Let $w \in L^\infty([0, T] \times \mathbb{R}^m)$ and $g \in L^\infty([0, T] \times \mathbb{R}^m, \mathbb{R}^m)$ be such that

$$\begin{cases} \partial_t w + D_x \cdot g \leq 0, \\ w(0, \cdot) = \bar{w}, \end{cases} \quad (44)$$

and $|g| \leq Cw$. Then, for $\mathcal{L}^1$ -a.e. $\tau \in]0, T]$, we have that

$$\int_{B_R(x_0)} w(\tau, \cdot) dx \leq \int_{B_{R+C\tau}(x_0)} \bar{w}(x) dx \quad \text{for every } x_0 \in \mathbb{R}^n \text{ and } R > 0. \quad (45)$$

PROOF. Let $\tau \in]0, T]$ be such that

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_{\tau-\varepsilon}^{\tau+\varepsilon} \int_K |w(t, x) - w(\tau, x)| dx dt = 0 \quad (46)$$

for every compact set $K \subset \mathbb{R}^m$. We will prove the statement of the lemma for any such τ .

Without loss of generality we assume $x_0 = 0$. Let $\chi_\varepsilon \in C^\infty(\mathbb{R}^+)$ be such that

$$\chi_\varepsilon = 1 \quad \text{on } [0, 1], \quad \chi_\varepsilon = 0 \quad \text{on } [1 + \varepsilon, +\infty[\quad \text{and} \quad \chi'_\varepsilon \leq 0.$$

Define the test function $\varphi(t, x) := \chi_\varepsilon(|x|/(R + C(\tau - t)))$. Note that φ is nonnegative and belongs to $C^\infty([0, \tau] \times \mathbb{R}^m)$. Note that we can test (44) with $\varphi(t, x) \mathbb{1}_{[-1, \tau]}(t)$. Indeed let μ be the measure $\partial_t w + D_x \cdot g$. Consider a standard family of nonnegative mollifiers $\xi^\delta \in C^\infty(\mathbb{R})$ and set $\zeta^\delta := \mathbb{1}_{[-1, \tau]} * \xi^\delta$. Testing (44) with $\varphi(t, x) \zeta^\delta(t)$ we get

$$\begin{aligned} & \int w(s, y) \varphi(s, y) \xi^\delta(\tau - s) ds dy - \int_{\mathbb{R}^m} \bar{w}(y) \varphi(0, y) dy \\ &= \int \zeta^\delta [w \partial_t \varphi + g \cdot \nabla_x \varphi] + \int \zeta^\delta \varphi d\mu. \end{aligned} \quad (47)$$

Note that $\int \zeta^\delta d\mu \leq 0$. Moreover, by (46), the integral

$$\int w(s, y) \varphi(s, y) \xi^\delta(\tau - s) ds dy$$

converge to $\int \varphi(\tau, x) w(\tau, x) dx$ as $\delta \downarrow 0$. Hence, in the limit we get

$$\begin{aligned} & \int_{[0, \tau] \times \mathbb{R}^n} [w \partial_t \varphi + g \cdot \nabla_x \varphi] \\ & \geq \int_{\mathbb{R}^n} \varphi(\tau, x) w(\tau, x) dx - \int_{\mathbb{R}^n} \varphi(0, x) \bar{w}(x) dx. \end{aligned} \quad (48)$$

We compute $w(s, y) \partial_t \varphi(s, y) + g(s, y) \cdot \nabla_x \varphi(s, y)$ as

$$\chi'_\varepsilon \left(\frac{|y|}{R + C(\tau - s)} \right) \left[\frac{C|y|w(s, y)}{(R + C(\tau - s))^2} + \frac{y \cdot g(s, y)}{|y|(R + C(\tau - s))} \right]. \quad (49)$$

Letting $\alpha := |y|/((R + C(\tau - s)))$, the expression in (49) becomes

$$\frac{\chi'_\varepsilon(\alpha)}{R + C(\tau - s)} \left[Cw\alpha + g \cdot \frac{y}{|y|} \right].$$

For $\alpha \leq 1$ we have $\chi'_\varepsilon(\alpha) = 0$, whereas for $\alpha \geq 1$ we have $\chi'_\varepsilon(\alpha) \leq 0$ and $Cw\alpha \geq |g|$. Thus we conclude that the integrand of the left-hand side of (48) is nonpositive. Hence

$$\int_{\mathbb{R}^m} \chi_\varepsilon \left(\frac{|x|}{R} \right) w(\tau, y) \, dx \leq \int_{\mathbb{R}^m} \chi_\varepsilon \left(\frac{|x|}{R + C\tau} \right) \bar{w}(y) \, dy.$$

Letting $\varepsilon \downarrow 0$ we get (45). □

PROOF OF PROPOSITION 3.13.

Existence. Let $\bar{u} \in L^\infty(\mathbb{R}^m)$ be given and consider a standard family of mollifiers $\{\eta_\varepsilon\}$ in $\mathbb{R}^m$ and a standard family of mollifiers ζ_ε in $\mathbb{R}$, the latter generated by a kernel $\zeta \in C_c^\infty([0, \infty[)$. Then consider the functions $\rho_\varepsilon \in C^\infty([0, \infty[\times \mathbb{R}^m)$ and $b_\varepsilon \in C^\infty([0, \infty[\times \mathbb{R}^m, \mathbb{R}^m)$ given by

$$\bar{u}_\varepsilon := \bar{u} * \eta_\varepsilon, \quad \rho_\varepsilon := \varepsilon + \rho * (\eta_\varepsilon \zeta_\varepsilon) \quad \text{and} \quad b_\varepsilon := \frac{(b\rho) * (\eta_\varepsilon \zeta_\varepsilon)}{\rho_\varepsilon}.$$

Note that

- (i) b_ε is Lipschitz for every ε ;
- (ii) $\|b_\varepsilon\|_\infty + \|\rho_\varepsilon\|_\infty + \|\bar{u}_\varepsilon\|_\infty$ is uniformly bounded;
- (iii) $b_\varepsilon \rightarrow b$ and $\rho_\varepsilon \rightarrow \rho$ strongly in L^1_{loc} ;
- (iv) $\partial_t \rho_\varepsilon + D_x \cdot (\rho_\varepsilon b_\varepsilon) = 0$ in the classical sense;
- (v) $\rho_\varepsilon(0, \cdot)$ converges weakly* in L^∞ to $\bar{\rho}$, see Lemma 3.7 and Remark 3.8.

Since b_ε is Lipschitz we can solve globally in time

$$\begin{cases} \partial_t \Phi_\varepsilon(t, x) = b_\varepsilon(t, \Phi_\varepsilon(t, x)), \\ \Phi_\varepsilon(0, x) = x. \end{cases}$$

Each $\Phi_\varepsilon(t, \cdot)$ is a diffeomorphism of $\mathbb{R}^m$. Thus, $u_\varepsilon(t, x) := \bar{u}([\Phi_\varepsilon(t, \cdot)]^{-1}(x))$ solves the equation

$$\begin{cases} \partial_t u_\varepsilon + b_\varepsilon \cdot \nabla_x u_\varepsilon = 0, \\ u_\varepsilon(0, \cdot) = \bar{u}_\varepsilon. \end{cases}$$

Using the chain rule and (iv) we conclude that

$$\begin{cases} \partial_t (u_\varepsilon \rho_\varepsilon) + D_x \cdot (\rho_\varepsilon b_\varepsilon u_\varepsilon) = 0, \\ [\rho_\varepsilon u_\varepsilon](0, \cdot) = \rho_\varepsilon(0, \cdot) \bar{u}_\varepsilon. \end{cases} \quad (50)$$

Due to (ii) we can extract a subsequence $\varepsilon_n \downarrow 0$ such that u_{ε_n} converges weakly* in L^∞ to some $u \in L^\infty$. From (ii), (iii) and (v), we conclude that:

- $u_{\varepsilon_n} \rho_{\varepsilon_n} \rightharpoonup^* u \rho$ and $b_{\varepsilon_n} \rho_{\varepsilon_n} u_{\varepsilon_n} \rightharpoonup^* b \rho u$ in $L^\infty([0, \infty[\times \mathbb{R}^m)$;
- $u_{\varepsilon_n} \rho_{\varepsilon_n}(0, \cdot) \rightharpoonup^* \bar{u} \bar{\rho}$ in $L^\infty(\mathbb{R}^m)$.

Passing into the limit in the distributional formulation of (50) we conclude that u solves (40) in the sense of distributions.

Comparison principle. Let u_i and $\bar{u}_i$ be given as in the statement of the second part of the proposition. We apply the renormalization property to $v := (u_2 - u_1)_+$ to get

$$\begin{cases} \partial_t(\rho v) + D_x \cdot (\rho v b) = 0, \\ [v \rho](0, \cdot) = 0. \end{cases} \quad (51)$$

Then we apply Lemma 3.17 with $w = \rho v$ and $g = \rho v b$ and we conclude that for $\mathcal{L}^1$ -a.e. t we have

$$\int_{\mathbb{R}^n} \rho(t, x) v(t, x) dx = 0.$$

Since $v \geq 0$ and $\rho \geq 0$, we conclude $\rho v = 0$, and hence $\rho u_1 \geq \rho u_2$. $\square$

PROOF OF COROLLARY 3.14. The existence has been proved in the previous proposition. Moreover, from the comparison principle proved above, the uniqueness of solutions of (40) for b and ρ as in the statement readily follows.

Next, recalling Lemma 3.7, up to changing their value on a set of measure zero, we have that $t \mapsto \rho(t, \cdot)$ and $t \mapsto \rho(t, \cdot) u(t, \cdot)$ are weakly* continuous. Consider $\zeta = \rho u^2$. Similarly, we conclude from Lemma 3.7 that there exists a $\hat{\zeta}$ such that $\hat{\zeta} = \zeta$ a.e. and $t \mapsto \hat{\zeta}(t, \cdot)$ is weakly* continuous. Therefore, for every $T > 0$, $\hat{\zeta}$ solves

$$\begin{cases} \partial_t \hat{\zeta} + D_x \cdot (\hat{\zeta} b) = 0, \\ \hat{\zeta}(0, \cdot) = \hat{\zeta}(0, \cdot), \\ \hat{\zeta}(T, \cdot) = \hat{\zeta}(T, \cdot), \end{cases}$$

in the sense of distributions. On the other hand, from the renormalization property we have

$$\begin{cases} \partial_t \hat{\zeta} + D_x \cdot (\hat{\zeta} b) = 0, \\ \hat{\zeta}(0, \cdot) = \rho(0, \cdot) [u(0, \cdot)]^2, \\ \hat{\zeta}(T, \cdot) = \rho(T, \cdot) [u(T, \cdot)]^2. \end{cases}$$

Thus, we conclude that $\rho(T, \cdot) [u(T, \cdot)]^2 = \hat{\zeta}(T, \cdot)$ for every T and hence $t \mapsto \rho(t, \cdot) \times [u(t, \cdot)]^2$ is weakly* continuous. For any $\tau \geq 0$ consider

$$\begin{aligned} & \rho(\tau, \cdot) (u(t, \cdot) - u(\tau, \cdot))^2 \\ &= \rho(\tau, \cdot) [u(t, \cdot)]^2 - 2[\rho(\tau, \cdot) u(\tau, \cdot)] u(t, \cdot) + \rho(\tau, \cdot) [u(\tau, \cdot)]^2. \end{aligned}$$

It follows that, for $\tau \rightarrow t$, $\rho(\tau, \cdot)(u(t, \cdot) - u(\tau, \cdot))^2 \rightharpoonup^* 0$ in L^∞ . Since $\rho(\tau, \cdot) \geq C > 0$ for every τ , we conclude that $u(\tau, \cdot) \rightarrow u(t, \cdot)$ strongly in L^1_{loc} . This proves that $u \mapsto u(t, \cdot)$ is strongly continuous in L^1_{loc} . $\square$

Corollary 3.15 follows trivially from Proposition 3.13.

REMARK 3.18. Clearly, the proof of the previous proposition can be used to solve transport and continuity equations in both directions and starting from any time T . Namely, under the same assumptions, for every $T \in \mathbb{R}$ and every bounded $\hat{u}$ and $\bar{u}$ there exist unique solutions to both the forward and the backward transport equations:

$$\begin{cases} \partial_t(\rho u) + D_x \cdot (\rho u b) = 0 & \text{in }]\infty, T] \times \mathbb{R}^n, \\ [\rho u](T, \cdot) = \rho(T, \cdot) \hat{u}, \end{cases} \quad (52)$$

$$\begin{cases} \partial_t(\rho u) + D_x \cdot (\rho u b) = 0 & \text{in } [T, \infty[\times \mathbb{R}^n, \\ [\rho u](T, \cdot) = \rho(T, \cdot) \bar{u}. \end{cases} \quad (53)$$

3.4. Stability of solutions to transport equations

The uniqueness results proved in the previous section have the following easy corollary.

COROLLARY 3.19. *Let $\{b_n\} \subset L^\infty([0, \infty[\times \mathbb{R}^m)$ be a sequence of vector fields converging strongly in L^1_{loc} to a bounded nearly incompressible vector field b with the renormalization property. Let ζ_n be solutions of*

$$\begin{cases} \partial_t \zeta_n + D_x \cdot (\zeta_n b_n) = 0, \\ \zeta_n(0, \cdot) = \bar{\zeta}_n. \end{cases} \quad (54)$$

If $\|\zeta_n\|_\infty$ is uniformly bounded and $\bar{\zeta}_n \rightharpoonup^ \bar{\zeta}$ in L^∞ , then ζ_n converges weakly* in L^∞ to the unique solution ζ of*

$$\begin{cases} \partial_t \zeta + D_x \cdot (\zeta b) = 0, \\ \zeta(0, \cdot) = \bar{\zeta}. \end{cases} \quad (55)$$

PROOF. If $\tilde{\zeta}$ is the weak* limit of any subsequence of $\{\zeta_n\}$, then $\tilde{\zeta}$ solves (55). Since the solution to such equation is unique, it follows that the whole sequence converges weakly* to ζ . $\square$

COROLLARY 3.20. *Let $\{b_n\}$, $b \in L^\infty([0, \infty[\times \mathbb{R}^m, \mathbb{R}^m)$, $\{\zeta_n\}$, ζ , $\{u_n\}$, $u \in L^\infty([0, \infty[\times \mathbb{R}^m)$ and $\bar{\rho}_n, \bar{\rho}, \bar{u}_n, \bar{u} \in L^\infty(\mathbb{R}^m)$ be such that*

- (a) $\zeta, \zeta_n > 0$, $\zeta^{-1}, \zeta_n^{-1} \in L^\infty$ and $\|\zeta_n\|_\infty + \|\zeta_n^{-1}\|_\infty + \|\bar{u}_n\|_\infty$ is uniformly bounded;
- (b) $\{b_n\}$ and b have the renormalization property and $b_n \rightarrow b$ in L^1_{loc} ;
- (c) $\partial_t \zeta + D_x \cdot (\zeta b) = \partial_t \zeta_n + D_x \cdot (\zeta_n b_n) = 0$;

(d) u_n and u solve

$$\begin{cases} \partial_t(\zeta_n u_n) + D_x \cdot (\zeta_n u_n b_n) = 0, \\ [\zeta_n u_n](0, \cdot) = \zeta_n(0, \cdot) \bar{u}_n, \end{cases} \quad (56)$$

$$\begin{cases} \partial_t(\zeta u) + D_x \cdot (\zeta u b) = 0, \\ [\zeta u](0, \cdot) = \zeta(0, \cdot) \bar{u}. \end{cases} \quad (57)$$

If $\zeta_n(0, \cdot) \rightharpoonup^* \zeta(0, \cdot)$ in L^∞ and $\bar{u}_n \rightarrow \bar{u}$ in L^1_{loc} , then $u_n \rightarrow u$ in L^1_{loc} .

PROOF. From the comparison principle of Proposition 3.13 it follows that $\|u_n\|_\infty \leq \|\bar{u}_n\|_\infty$. Moreover, from Corollary 3.19 it follows that $\zeta_n \rightharpoonup^* \zeta$.

Set $\beta_n := \zeta_n u_n$ and $\bar{\beta}_n := \zeta_n(0, \cdot) \bar{u}_n$. We conclude from Corollary 3.19 that β_n converges weakly* in L^∞ to the unique solution β of

$$\begin{cases} \partial_t \beta + D_x \cdot (\beta b) = 0, \\ \beta(0, \cdot) = \zeta(0, \cdot) \bar{u}. \end{cases} \quad (58)$$

Therefore, by Corollary 3.14, $\beta/\zeta = u$. Applying the renormalization property, we conclude that $v_n := u_n^2$ and $v := u^2$ solve

$$\begin{cases} \partial_t(\zeta_n v_n) + D_x \cdot (\zeta_n v_n b_n) = 0, \\ [\zeta_n v_n](0, \cdot) = \zeta_n(0, \cdot) \bar{u}_n^2, \end{cases} \quad (59)$$

$$\begin{cases} \partial_t(\zeta v) + D_x \cdot (\zeta v b) = 0, \\ [\zeta v](0, \cdot) = \zeta(0, \cdot) \bar{u}^2. \end{cases} \quad (60)$$

Therefore, applying the argument above we conclude that $\zeta_n u_n^2 \rightharpoonup^* \zeta u^2$. Note that

$$\zeta_n (u_n - u)^2 = \zeta_n u_n^2 + \zeta_n u^2 - 2\zeta_n u_n u \rightharpoonup^* \zeta u^2 + \zeta u^2 - 2\zeta u u = 0.$$

Since for some constant C we have $\zeta_n \geq C$ for every n , we conclude that $(u_n - u)^2 \rightarrow 0$ strongly in L^1_{loc} . $\square$

In the same way we can prove the following more refined version of the previous corollary, which will be used in studying the well-posedness for the Keyfitz and Kranzer system.

COROLLARY 3.21. Assume that

- the pairs $\{(b_n, \rho_n)\}_n, (b, \rho)$ have the renormalization property and $\rho_n \geq 0$;
- $(b_n, \rho_n) \rightarrow (b, \rho)$ in L^1_{loc} and $\|b_n\|_\infty + \|\rho_n\|_\infty$ is uniformly bounded;
- the traces $\rho_n(0, \cdot) \rightarrow \rho(0, \cdot)$ and $\bar{u}_n \rightarrow \bar{u}$ strongly in L^1_{loc} .

If u_n, u solve (56) and (57), then $\rho_n u_n \rightarrow \rho u$ strongly in L^1_{loc} .

PROOF. From the proof of Corollary 3.20 we conclude that $\rho_n (u_n - u)^2 \rightarrow 0$ strongly in L^1_{loc} . Since $\|\rho_n\|_\infty$ is uniformly bounded, we get that $(\rho_n u_n - \rho_n u)^2 \rightarrow 0$, and hence

$|\rho_n u_n - \rho_n u| \rightarrow 0$ strongly in L^1_{loc} . But $|u\rho_n - \rho u| \leq \|u\|_\infty |\rho_n - \rho| \rightarrow 0$ strongly in L^1_{loc} , and thus we finally get $|\rho_n u_n - \rho u| \rightarrow 0$, which is the desired conclusion. $\square$

3.5. Existence, uniqueness and stability of regular Lagrangian flows

We will now show existence, uniqueness, and stability of the regular Lagrangian flows using the stability results for transport and continuity equations proved in the previous sections.

THEOREM 3.22. *Let b be a bounded nearly incompressible vector field with the renormalization property. Then there exists a unique regular Lagrangian flow Φ for b . Moreover, let b_n be a sequence of bounded nearly incompressible vector fields with the renormalization property such that*

- $\|b_n\|_\infty$ is uniformly bounded and $b_n \rightarrow b$ strongly in L^1_{loc} ;
- the densities ρ_n generated by b_n satisfy $\limsup_n (\|\rho_n\|_\infty + \|\rho_n^{-1}\|_\infty) < \infty$.

Then the regular Lagrangian flows Φ_n generated by b_n converge in L^1_{loc} to Φ .

PROOF.

Uniqueness. Let Φ and Ψ be two regular Lagrangian flows associated to the same nearly incompressible vector field. For any $\zeta \in L^\infty(\mathbb{R}^n)$ consider the bounded functions ζ and $\hat{\zeta}$ given by

$$\begin{aligned} \int \varphi(t, x) \zeta(t, x) dt dx &= \int \varphi(t, \Phi(t, x)) \bar{\zeta}(x) dt dx, \\ \int \varphi(t, x) \hat{\zeta}(t, x) dt dx &= \int \varphi(t, \Psi(t, x)) \bar{\zeta}(x) dt dx. \end{aligned}$$

According to Proposition 3.5, ζ and $\hat{\zeta}$ solve both the same equation

$$\begin{cases} \partial_t \zeta + D_x \cdot (\zeta b) = 0, \\ \zeta(0, \cdot) = \bar{\zeta}. \end{cases}$$

When b has the renormalization property we can apply Proposition 3.13 to conclude that $\zeta = \hat{\zeta}$. Therefore, when b has the renormalization property we conclude that, for any compactly supported $\varphi \in L^\infty(\mathbb{R} \times \mathbb{R}^m)$ and $\bar{\zeta} \in L^\infty(\mathbb{R}^m)$, we have

$$\int \varphi(t, \Phi(t, x)) \bar{\zeta}(x) dt dx = \int \varphi(t, \Psi(t, x)) \bar{\zeta}(x) dt dx.$$

This easily implies that $\Psi = \Phi$ $\mathcal{L}^{m+1}$ -a.e.

Stability. Next consider a sequence of $b_n \rightarrow b$ as in the statement of the proposition. Let Φ and Φ_n be regular Lagrangian flows generated by b and b_n . Fix again any $\bar{\zeta} \in L^\infty$ and define ζ as in the previous step and ζ_n by

$$\int \varphi(t, x) \zeta_n(t, x) = \int \varphi(t, \Phi_n(t, x)) \bar{\zeta}(x) dt dx.$$

Applying the comparison principle we get that $\|\zeta_n\|_\infty$ is uniformly bounded, and from Corollary 3.19 we conclude that $\zeta_n \rightharpoonup^* \zeta$. Therefore we get that

$$\int \varphi(t, \Phi_n(t, x)) \bar{\zeta}(x) dt dx \rightarrow \int \varphi(t, \Phi(t, x)) \bar{\zeta}(x) dt dx \quad (61)$$

for every bounded $\bar{\zeta}$ and every φ which is bounded and has bounded support.

Note that, since $\|b_n\|_\infty$ is uniformly bounded, for every $R > 0$, $\|\Phi_n\|_{L^\infty([0, R] \times B_R(0))}$ is uniformly bounded. Therefore, if $\bar{\zeta}$ has bounded support, then (61) holds for every bounded φ which has support bounded in time. Thus, we can apply (61) with $\bar{\zeta} = \mathbb{1}_{B_R(0)}$ and $\varphi(t, x) = \mathbb{1}_{[0, R]}(t) |x|^2$ in order to get

$$\int_{[0, R] \times B_R(0)} |\Phi_n(t, x)|^2 dt dx \rightarrow \int_{[0, R] \times B_R(0)} |\Phi(t, x)|^2 dt dx. \quad (62)$$

Next, apply (61) with $\varphi(t, x) = \mathbb{1}_{[0, R]}(t) \gamma(t) x \cdot v$ and $\bar{\zeta} = \beta \mathbb{1}_{B_R(0)}$. Then we conclude that

$$\begin{aligned} & \int_{[0, R] \times B_R(0)} \Phi_n(t, x) \cdot v \gamma(t) \beta(x) dt dx \\ & \rightarrow \int_{[0, R] \times B_R(0)} \Phi(t, x) \cdot v \gamma(t) \beta(x) dt dx. \end{aligned}$$

By linearity, we conclude that

$$\begin{aligned} & \int_{[0, R] \times B_R(0)} \sum_{i=1}^N \Phi_n(t, x) \cdot v_i \gamma_i(t) \beta_i(x) dt dx \\ & \rightarrow \int_{[0, R] \times B_R(0)} \sum_{i=1}^N \Phi(t, x) \cdot v_i \gamma_i(t) \beta_i(x) dt dx \end{aligned}$$

for any choice of the bounded functions γ_i , β_i and v_i . However, by a standard argument, we can approximate Φ strongly in $L^1([0, R] \times B_R(0))$ with functions of type $\sum_{i=1}^N v_i \gamma_i(t) \beta_i(x)$. This gives

$$\int_{[0, R] \times B_R(0)} \Phi_n(t, x) \cdot \Phi(t, x) dt dx \rightarrow \int_{[0, R] \times B_R(0)} |\Phi(t, x)|^2 dt dx. \quad (63)$$

Therefore, from (62) and (63), we get

$$\lim_{n \uparrow \infty} \int_{[0, R] \times B_R(0)} |\Phi_n(t, x) - \Phi(t, x)|^2 dt dx = 0.$$

From the arbitrariness of R we conclude that $\Phi_n \rightarrow \Phi$ in L^1_{loc} .

Existence. Step 1: Regular approximation. We finally address the existence of a regular Lagrangian flow. Fix two kernels $\chi \in C_c^\infty([0, \infty[)$ and $\psi \in C^\infty(\mathbb{R}^m)$, let $\{\chi_\varepsilon\}_\varepsilon$ and $\{\psi_\varepsilon\}_\varepsilon$ be the two standard families of mollifiers generated by χ and η , and set $\varphi_\varepsilon(t, x) := \chi_\varepsilon(t) \eta_\varepsilon(x)$.

Let ρ be the density generated by b and set $\rho_\varepsilon := \rho * \varphi_\varepsilon$, $b_\varepsilon := b * \varphi_\varepsilon / \rho_\varepsilon$. Note that

- $\|b_\varepsilon\|_\infty + \|\rho_\varepsilon\|_\infty + \|\rho_\varepsilon^{-1}\|_\infty$ is uniformly bounded;
- $b_\varepsilon \rightarrow b$ and $\rho_\varepsilon \rightarrow \rho$ in L^1_{loc} ;
- $\rho_\varepsilon(t, \cdot) \rightharpoonup^* \rho(t, \cdot)$ in $L^\infty(\mathbb{R}^m)$ for every $t \geq 0$.

For each ε , b_ε is globally Lipschitz, and therefore we can apply the classical Cauchy–Lipschitz theorem to get the unique regular Lagrangian flow Φ_ε generated by b_ε .

Note that $\|\Phi_\varepsilon\|_{L^\infty(K)}$ is uniformly bounded for every compact set K . Thus we can extract a sequence $\{\Phi_n\} = \{\Phi_{\varepsilon_n}\}$ which locally converges weakly* to a map Φ . We will show that Φ_n converges strongly in L^1_{loc} . From this we easily conclude that Φ is a regular Lagrangian flow for b . From now on, in order to simplify the notation we will use b_n , ρ_n for b_{ε_n} and ρ_{ε_n} .

Existence. Step 2: Strong convergence. Note that each $\Phi_n(t, \cdot)$ is a diffeomorphism of $\mathbb{R}^m$. Therefore we can define $\Psi_n(t, \cdot) := [\Phi_n(t, \cdot)]^{-1}$. Fix $T > 0$ and solve the following ODE backward in time:

$$\begin{cases} \frac{d}{dt} \Lambda_n(t, x) = b_n(t, \Lambda_n(t, x)), \\ \Lambda_n(T, x) = x. \end{cases}$$

Note that $\Lambda_n(t, \cdot) = \Phi_n(t, \Psi_n(T, \cdot))$. Thus, if we denote by $J_n(t, \cdot)$ the Jacobian of $\Lambda_n(t, \cdot)$, we get that $0 \leq C^{-2} \leq J_n(t, \cdot) \leq C^2$. Denote by $\Gamma_n(t, \cdot)$ the inverse of $\Lambda_n(t, \cdot)$ and set $\zeta_n(t, x) := J_n(t, \Gamma_n(t, x))$. Moreover, for every $\bar{w} \in L^\infty(\mathbb{R}^m, \mathbb{R}^m)$ define the function $w_n(t, x) := \bar{w}(\Gamma_n(t, x))$. Clearly we have

$$\begin{cases} \partial_t(\zeta_n w_n) + D_x \cdot (\zeta_n w_n \otimes b_n) = 0 & \text{on } [0, T] \times \mathbb{R}^m, \\ \zeta_n w_n(T, x) = \bar{w}(x). \end{cases}$$

(The first line is just a shorthand notation for the equations $\partial_t(\zeta_n w_n^i) + D_x \cdot (\zeta_n w_n^i b_n) = 0$ for $i \in \{1, \dots, m\}$.) We claim that the ζ_n 's have a unique weak* limit. Indeed, assume that ζ and $\hat{\zeta}$ are weak* limits of two convergent subsequences of ζ_n 's. Then $\partial_t \zeta + D_x \cdot (b \zeta) = 0$ and $\partial_t \hat{\zeta} + D_x \cdot (b \hat{\zeta}) = 0$. Moreover, both ζ and $\hat{\zeta}$ have weak trace equal to 1 at $t = T$. Thus by the backward uniqueness of Remark 3.18, we conclude that ζ and $\hat{\zeta}$ coincide with the unique solution of

$$\begin{cases} \partial_t \beta + D_x \cdot (\beta b) = 0 & \text{on } [0, T] \times \mathbb{R}^n, \\ \beta(T, \cdot) = 1. \end{cases}$$

Note that there exists a constant C such that $|\Gamma_n(t, x) - x| \leq C(T - t)$ for every t, x and j . Fix $r > 0$ and choose $R > 0$ so large that $R - CT > r$. Let $\bar{w}$ be the vector-valued map $x \rightarrow x \mathbb{1}_{B_R(0)}(x)$. Thus, for every $t < T$ and every $|x| < r$, $w_n(t, x)$ is equal to the vector $\Gamma_n(t, x)$. Thanks to Remark 3.18, w_n converges strongly in L^1_{loc} the unique w solving

$$\begin{cases} \partial_t(\beta w) + D_x \cdot (\beta w \otimes b) = 0 & \text{on } [0, T] \times \mathbb{R}^m, \\ [\beta w](0, \cdot) = \bar{w}. \end{cases}$$

Hence, by the arbitrariness of r we conclude that Γ_n converges to a unique Γ strongly in L^1_{loc} .

For each x , $\Gamma_n(\cdot, x)$ is a Lipschitz curve, with Lipschitz constant uniformly bounded. Thus we infer that, for a.e. x , $\Gamma_n(\cdot, x)$ converges uniformly to the curve $\Gamma(\cdot, x)$ on $[0, T]$. Hence, we conclude that, after possibly changing Γ on a set of measure 0, for every $t \geq 0$ the maps $\Gamma_n(t, \cdot)$ converge to $\Gamma(t, \cdot)$ in $L^1_{\text{loc}}(\mathbb{R}^m)$.

Since $\Gamma_n(0, \cdot) = \Phi_n(T, \cdot)$ we conclude that for every T there exists a $\Phi(T, \cdot)$ such that $\Phi_n(T, \cdot)$ converges to $\Phi(T, \cdot)$ in $L^1_{\text{loc}}(\mathbb{R}^m)$. Since Φ_n is locally uniformly bounded, we conclude that Φ_n converges to Φ strongly in $L^1_{\text{loc}}(\mathbb{R}^+ \times \mathbb{R}^m)$.

Existence. Step 3: Near incompressibility. Note that, by our construction, there exists a constant C such that, for every t and every n ,

$$C^{-1} \mathcal{L}^m \leq \Phi_n(t, \cdot)_{\#} \mathcal{L}^m \leq C \mathcal{L}^m. \quad (64)$$

Let $\varphi \in C_c([0, \infty[\times \mathbb{R}^m)$ be given. Then

$$\int |\varphi(t, \Phi_n(t, x))| dx dt \leq C \int |\varphi(t, y)| dy dt < \infty.$$

Up to extracting another subsequence, not relabeled, we can assume that $\Phi_n(t, x) \rightarrow \Phi(t, x)$ for $\mathcal{L}^{m+1}$ -a.e. (t, x) . Thus, by the dominated convergence theorem:

$$\begin{aligned} \lim_{n \uparrow \infty} \int \varphi d\mu_{\Phi_n} &= \lim_{n \uparrow \infty} \int \varphi(t, \Phi_n(t, x)) dx dt \\ &= \int \varphi(t, \Phi(t, x)) dt dx \\ &= \int \varphi d\mu_{\Phi}. \end{aligned}$$

Therefore, from (64) we get $C^{-1} \mathcal{L}^{m+1} \leq \mu_{\Phi} \leq C \mathcal{L}^{m+1}$. Therefore Φ satisfies condition (a) of Definition 3.1.

Existence. Step 4: Final ODE. Next, we show that $b_n(t, \Phi_n(t, x)) \rightarrow b(t, \Phi(t, x))$ strongly in L^1_{loc} , from which (b) of Definition 3.1 follows. Let R be any given positive number. Since $\|b_n\|_{\infty} \leq C$, we have $\|\Phi_n\|_{L^{\infty}([0, R] \times B_R(0))} \leq (C + 1)R$. Thus, set

$b'_n := b_n \mathbb{1}_{[0,R] \times B_{(C+1)R}(0)}$ and $b' := b \mathbb{1}_{[0,R] \times B_{(C+1)R}(0)}$. Using Egorov's and Lusin's theorems, for any given $\varepsilon > 0$ choose $\hat{b}_n, \hat{b} \in C_c([0, \infty[\times \mathbb{R}^m)$ such that

- $\|\hat{b}_n - b'_n\|_{L^1} + \|\hat{b} - b'\|_{L^1} < \varepsilon$;
- $\hat{b}_n \rightarrow \hat{b}$ uniformly.

Then, $\hat{b}_n(t, \Phi_n(t, x)) \rightarrow \hat{b}(t, \Phi(t, x))$ for $\mathcal{L}^{m+1}$ -a.e. (t, x) . Thus,

$$\begin{aligned}
 & \limsup_{n \uparrow \infty} \|b_n(\cdot, \Phi_n(\cdot)) - b(\cdot, \Phi(\cdot))\|_{L^1([0,R] \times B_R(0))} \\
 &= \limsup_{n \uparrow \infty} \|b'_n(\cdot, \Phi_n(\cdot)) - b'(\cdot, \Phi(\cdot))\|_{L^1([0,R] \times B_R(0))} \\
 &\leq \limsup_{n \uparrow \infty} \|\hat{b}_n(\cdot, \Phi_n(\cdot)) - \hat{b}(\cdot, \Phi(\cdot))\|_{L^1([0,R] \times B_R(0))} \\
 &\quad + \limsup_{n \uparrow \infty} (\|\hat{b}_n - b'_n\|_{L^1} + \|\hat{b} - b'\|_{L^1}) \\
 &= \limsup_{n \uparrow \infty} (\|\hat{b}_n - b'_n\|_{L^1} + \|\hat{b} - b'\|_{L^1}) \\
 &\stackrel{(64)}{\leq} C \limsup_{n \uparrow \infty} (\|\hat{b}_n - b'_n\|_{L^1} + \|\hat{b} - b'\|_{L^1}) \\
 &\leq C\varepsilon.
 \end{aligned}$$

By the arbitrariness of R and ε , we get the desired convergence. This completes the proof. $\square$

4. Commutator estimates and Ambrosio's renormalization theorem

In this section we study the following problem. Let $\Omega \subset \mathbb{R}^d$ be an open set and $B : \Omega \rightarrow \mathbb{R}^d$ a bounded BV vector field. Assume $w^1, \dots, w^k$ are L^∞ functions which satisfy

$$D \cdot (w^i B) = 0 \quad \text{distributionally in } \Omega \text{ for every } i$$

(that is, $D \cdot (w \otimes B) = 0$) and let $H \in C^1(\mathbb{R}^k)$. What are the properties of the distribution $D \cdot (H(w)B)$?

In particular, our final goal is to show the following theorem, which has been proved in [10] by slightly adapting the ideas of [2].

THEOREM 4.1. *Let B , Ω , w and H be as above. Then, $D \cdot (H(w)B)$ is a Radon measure and*

$$\left| D \cdot (H(w)B) - \left(H(w) - \sum_{i=1}^d \frac{\partial H}{\partial v_i}(w) w^i \right) D^a \cdot B \right| \leq C |D^s \cdot B|, \quad (65)$$

where the constant C depends only on $R := \|w\|_\infty$ and $\|H\|_{C^1(B_R(0))}$.

Our approach to this problem is to consider appropriate commutators and get estimates for them. More precisely, fix a standard kernel η in $\mathbb{R}^d$ supported in the ball $B_r(0)$ and let $\{\eta_\varepsilon\}_{\varepsilon>0}$ be the standard family of mollifiers generated by η . Thus, for any distribution T in Ω the convolution $T * \rho_\varepsilon$ is a well-defined distribution in the open set $\Omega_\varepsilon := \{x \in \Omega: \text{dist}(x, \partial\Omega) > \varepsilon r\}$. Since $w^i * \eta_\varepsilon \rightarrow w^i$ converges strongly in $L^1(K)$ to w^i for any $K \subset\subset \Omega$, we conclude $D \cdot (H(w * \eta_\delta)B)$ converges in the sense of distributions to $D \cdot (H(w)B)$ in every open set $\Omega' \subset\subset \Omega$. Since $w * \eta_\delta$ is smooth, the usual chain rule applies and we can compute

$$\begin{aligned} D \cdot (H(w * \eta_\delta)B) &= \sum_{i=1}^d \frac{\partial H}{\partial v_i}(w * \eta_\delta) D \cdot (w^i * \eta_\delta B) \\ &\quad + \left(H(w * \eta_\delta) - \sum_{i=1}^d \frac{\partial H}{\partial v_i}(w * \eta_\delta) w^i * \eta_\delta \right) D \cdot B. \end{aligned}$$

Moreover, notice that $(D \cdot (w^i B)) * \eta_\delta = 0$. Thus we can write

$$\begin{aligned} D \cdot (H(w * \eta_\delta)B) &= \sum_{i=1}^d \frac{\partial H}{\partial v_i}(w * \eta_\delta) [D \cdot (w^i * \eta_\delta B) - (D \cdot (w^i B)) * \eta_\delta] \\ &\quad + \left(H(w * \eta_\delta) - \sum_{i=1}^d \frac{\partial H}{\partial v_i}(w * \eta_\delta) w^i * \eta_\delta \right) D \cdot B. \end{aligned} \quad (66)$$

Motivated by these computations we introduce the following terminology and notation.

DEFINITION 4.2. For every fixed kernel η , we denote by $T_{\delta,\eta}^i$ the commutators

$$T_{\delta,\eta}^i := (D \cdot (B w^i)) * \eta_\delta - D \cdot (B w^i * \eta_\delta). \quad (67)$$

Moreover, the vector-valued distribution $(T_{\delta,\eta}^1, \dots, T_{\delta,\eta}^k)$ will be denoted by $T_{\delta,\eta}$. When no confusion can arise, we drop the η from $T_{\delta,\eta}^i$ and $T_{\delta,\eta}$.

Clearly, in our case the commutators $T_\delta = D \cdot (w \otimes B) * \eta_\delta - D \cdot ((w * \eta_\delta) \otimes B)$ are equal to $-D \cdot ((w * \eta_\delta) \otimes B)$. Since $w * \eta_\delta$ is smooth and B is a BV vector field, $(w * \eta_\delta) \otimes B$ is a BV matrix-valued function. Thus T_δ is a vector-valued measure. However, this turns out to hold even when we do not assume $D \cdot (w \otimes B) = 0$: The commutators T_δ are always measures, for every BV vector field B and every L^∞ map w (see Proposition 4.6(a)).

Next, write $D \cdot B = D^a \cdot B + D^s \cdot B$, and from (66) get the inequality

$$\begin{aligned} &\left| D \cdot (H(w * \eta_\delta)B) - \left(H(w * \eta_\delta) - \sum_{i=1}^d \frac{\partial H}{\partial v_i}(w * \eta_\delta) w^i * \eta_\delta \right) D^a \cdot B \right| \\ &\leq C(|T_{\delta,\eta}| + C|D^s \cdot B|), \end{aligned} \quad (68)$$

where the constant C depends on H and $\|w\|_\infty$.

Comparing (65) and (68), it is clear that we might try to prove Theorem 4.1 by careful analyzing the behavior of the commutators $|T_{\delta,\eta}|$. This is done in Proposition 4.6, with the help of a technical Proposition 4.3 concerning difference quotients of BV functions, which is proved in Section 4.1. The key commutator estimate of Proposition 4.6 is stated and proved in Section 4.2. In Section 4.3 we state two lemmas. The first one is due to Bouchut and it was used in the first proof of the results of [2], in combination with the rank-one theorem (see Theorem 2.13). The second lemma is a generalization of Bouchut's one, suggested by Alberti. This new lemma can replace the one by Bouchut and the rank-one theorem in the proof of Theorem 4.1, yielding a much more transparent and self-contained argument. In Section 4.4 we give both these proofs of Theorem 4.1.

4.1. Difference quotients of BV functions

In what follows, for BV vector fields B , we denote, as usual, by DB their distributional derivative, which are Radon measures. If $DB = M\mathcal{L}^d + D^s B$ is the Radon-Nikodym decomposition of DB with respect to $\mathcal{L}^d$, then we denote M by ∇B .

PROPOSITION 4.3. *Let $B \in BV_{\text{loc}}(\mathbb{R}^d, \mathbb{R}^m)$ and let $z \in \mathbb{R}^d$. Then the difference quotients*

$$\frac{B(x + \delta z) - B(x)}{\delta}$$

can be canonically written as $B_{1,\delta}(z)(x) + B_{2,\delta}(z)(x)$, where:

- (a) $B_{1,\delta}(z)$ converges strongly in L^1_{loc} to $\nabla B \cdot z$ as $\delta \downarrow 0$.
- (b) For any compact set $K \subset \mathbb{R}^d$, we have

$$\limsup_{\delta \downarrow 0} \int_K |B_{2,\delta}(z)(x)| dx \leq |D^s B \cdot z|(K). \quad (69)$$

- (c) For every compact set $K \subset \mathbb{R}^d$, we have

$$\sup_{\delta \in]0, \varepsilon[} \int_K |B_{1,\delta}(z)(x)| + |B_{2,\delta}(z)(x)| dx \leq |z| |DB|(K_\varepsilon), \quad (70)$$

where $K_\varepsilon := \{x: \text{dist}(x, K) \leq \varepsilon\}$.

REMARK 4.4. The decomposition of the proof is canonical in the sense that we give an explicit way of constructing $B_{1,\delta}$ and $B_{2,\delta}$ from the measures $D^a B \cdot z$ and $D^s B \cdot z$. One important consequence of this explicit construction is the following linearity property: If $B^1, B^2 \in BV_{\text{loc}}(\mathbb{R}^d, \mathbb{R}^m)$, $\lambda_1, \lambda_2 \in \mathbb{R}$ and $z \in \mathbb{R}^d$, then

$$(\lambda_1 B^1 + \lambda_2 B^2)_{i,\delta}(z)(x) = \lambda_1 B^1_{i,\delta}(z)(x) + \lambda_2 B^2_{i,\delta}(z)(x). \quad (71)$$

PROOF OF PROPOSITION 4.3. Let $e_1, \dots, e_d$ be orthonormal vectors in $\mathbb{R}^d$. In the corresponding system of coordinates we use the notation $x = (x_1, \dots, x_{d-1}, x_d) = (x', x_d)$. Without loss of generality we can assume that $z = e_d$. Recall the following elementary fact: If μ is a Radon measure on $\mathbb{R}$, then the functions

$$\hat{\mu}_\delta(t) := \frac{\mu([t, t + \delta])}{\delta} = \mu * \frac{\mathbb{1}_{[-\delta, 0]}}{\delta}(t), \quad t \in \mathbb{R},$$

satisfy

$$\int_K |\hat{\mu}_\delta| dt \leq \mu(K_\delta) \quad (72)$$

for every compact set $K \subset \mathbb{R}$, where K_δ denotes the δ -neighborhood of K .

Consider the measure $D_{e_d}B = DB \cdot e_d$, and the vector-valued function $\nabla B \cdot e_d$. Clearly this function is the Radon–Nikodym derivative of $D_{e_d}B$ with respect to $\mathcal{L}^d$ and we denote by $D_{e_d}^s B$ the singular measure $D^s B \cdot e_d = D_{e_d}B - \nabla B \cdot e_d \mathcal{L}^d$.

We define

$$B_{1,\delta}(x', x_d) = \frac{1}{\delta} \int_{x_d}^{x_d + \delta} \nabla B \cdot e_d(x', s) ds.$$

By Fubini–Tonelli's theorem and standard arguments on convolutions, we get that $B_{1,\delta} \rightarrow \nabla B \cdot e_d$ strongly in L^1_{loc} .

Next set

$$B_{2,\delta}(x', x_2) := \frac{B(x', x_d + \delta) - B(x', x_d)}{\delta} - B_{1,\delta}(x', x_d),$$

and, for $\mathcal{L}^{d-1}$ -a.e. $y \in \mathbb{R}^{d-1}$, define $B_y : \mathbb{R} \rightarrow \mathbb{R}$ by $B_y(s) = B(y, s)$.

We recall the following slicing properties of BV functions (see Theorems 3.103, 3.107 and 3.108 of [11]):

- (a) $B_y \in BV_{\text{loc}}(\mathbb{R}, \mathbb{R}^m)$ for $\mathcal{L}^{d-1}$ -a.e. y ;
- (b) if we let $D^s B_y + B'_y \mathcal{L}^1$ be the Radon–Nikodym decomposition of DB_y , then we have

$$\nabla B(y, s) \cdot e_d = B'_y(s) \quad \text{for } \mathcal{L}^d\text{-a.e. } (y, s)$$

and

$$|D_{e_d}^s|(A) = \int_{\mathbb{R}^{d-1}} |D^s B_y|(A \cap \{(y, s) : s \in \mathbb{R}\}) dy;$$

- (c) $B_y(s + \delta) - B_y(s) = DB_y([s, s + \delta])$.

Therefore, for any $\delta > 0$ and for $\mathcal{L}^{d-1}$ -a.e. y , we have

$$\begin{aligned} \frac{B(y, x_d + \delta) - B(y, x_d)}{\delta} &= \frac{B_y(x_d + \delta) - B_y(x_d)}{\delta} = \frac{DB_y([x_d, x_d + \delta])}{\delta} \\ &= (\widehat{B'_x \mathcal{L}^1})_\delta(x_d) + (\widehat{D^s B_y})_\delta(x_d) \\ &= B_{1,\delta}(y, x_d) + (\widehat{D^s B_y})_\delta(x_d) \quad \text{for } \mathcal{L}^1\text{-a.e. } x_d. \end{aligned}$$

Therefore

$$\begin{aligned} \int_K |B_{2,\delta}| &\leq \int_{\mathbb{R}^{d-1}} \int_{\{x_d: (y, x_d) \in K\}} |(\widehat{D^s B_y})_\delta(x_d)| \, dx_d \, dy \\ &\leq \int_{\mathbb{R}^{d-1}} |D^s B_y|(\{x_d: (y, x_d) \in K_\delta\}) \, dy \\ &= |D^s B \cdot e_d|(K_\delta) \leq |D^s B|(K_\delta). \end{aligned} \tag{73}$$

Letting $\delta \downarrow 0$, this gives (69).

Note, moreover, that

$$\begin{aligned} \int_K |B_{1,\delta}| &\leq \int_{\mathbb{R}^{d-1}} \int_{\{x_d: (y, x_d) \in K\}} |(\widehat{B'_y \mathcal{L}^1})_\delta(x_d)| \, dx_d \, dy \\ &\leq \int_{K_\delta} |\nabla B \cdot e_d|(y, x_d) \, dy \, dx_d \\ &\leq \int_{K_\delta} |\nabla B|(y, x_d) \, dy \, dx_d. \end{aligned} \tag{74}$$

Adding the bounds (73) and (74) we get (70). $\square$

4.2. Commutator estimate

In this subsection we use the technical proposition proved above in order to show the key commutator estimate which, together with Lemma 4.8 will give Theorem 4.1. In order to state it we introduce the following notation.

DEFINITION 4.5. For any $\eta \in C_c^\infty(\mathbb{R}^d)$ and any matrix M we define

$$\Lambda(M, \eta) := \int_{\mathbb{R}^d} |\nabla \eta(z) \cdot M \cdot z| \, dz. \tag{75}$$

PROPOSITION 4.6 (Commutators estimate). *Let $B \in BV \cap L^\infty(\Omega, \mathbb{R}^d)$ and $w \in L^\infty(\Omega, \mathbb{R}^k)$. Assume η is an even convolution kernel and denote by M the Borel matrix-valued measure given by the Radon–Nikodym decomposition $DB = M|DB|$. Then*

(a) the commutators (67) are induced by measures and the total variation of these measures is uniformly bounded on any compact subset of Ω ;

(b) any weak* limit σ of a subsequence of $\{|T_\delta|\}_{\delta \downarrow 0}$ as $\delta \downarrow 0$ is a singular measure which satisfies the bound

$$\sigma \llcorner A \leq \|w\|_{L^\infty(A)} (|D^s \cdot B| + \Lambda(M, \eta) |D^s B|) \quad \text{for any open set } A \subset \subset \Omega. \quad (76)$$

PROOF. Let $\delta > 0$ be fixed and choose $\lambda > 0$ such that the support of η is contained in $B_\lambda(0)$. Next, let A be any open set such that $\delta\lambda < \text{dist}(A, \partial\Omega)$. First of all, note that, in A , we have

$$T_\delta = r_\delta \mathcal{L}^d - w * \eta_\delta D \cdot B, \quad (77)$$

where r_δ is an L^1 function which will be computed below. Note that the formula $w * \eta_\delta D \cdot B$ makes sense, because $D \cdot B$ is a measure and $w * \eta_\delta$ is a continuous function.

Indeed, fix a test function $\varphi \in C_c^\infty(A)$ and notice that

$$\begin{aligned} \langle T_\delta^i, \varphi \rangle &= \langle D \cdot ((w^i B) * \eta_\delta), \varphi \rangle - \langle D \cdot (w^i * \eta_\delta B), \varphi \rangle \\ &= \int_{\mathbb{R}^d} D_x \cdot \left(\int_{\mathbb{R}^d} w(y) B(y) \eta_\delta(x - y) dy \right) \varphi(x) dx \\ &\quad + \int_{\mathbb{R}^d} w^i * \eta_\delta B \cdot \nabla \varphi \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} w(y) B(y) \cdot \nabla_x \eta_\delta(x - y) dy \varphi(x) dx \\ &\quad - \int_{\mathbb{R}^d} \nabla (w^i * \eta_\delta) \cdot B \varphi - \int_{\mathbb{R}^d} w^i * \eta_\delta \varphi d[D \cdot B] \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} w^i(y) B(y) \cdot \nabla_x \eta_\delta(x - y) dy \right) \varphi(x) dx \\ &\quad + \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} w^i(y) \nabla_y \eta_\delta(x - y) dy \right) \cdot B(x) \varphi(x) dx \\ &\quad - \int_{\mathbb{R}^d} w^i * \eta_\delta \varphi d[D \cdot B] \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} w^i(y) (B(x) - B(y)) \cdot \nabla_y \eta_\delta(x - y) dy \right) \varphi(x) dx \\ &\quad - \int_{\mathbb{R}^d} w^i * \eta_\delta \varphi d[D \cdot B]. \end{aligned}$$

This proves (77) with

$$\begin{aligned} r_\delta(x) &= \int_{\mathbb{R}^d} w(y) (B(x) - B(y)) \cdot \nabla_y \eta_\delta(x - y) \, dy \\ &= - \int_{\mathbb{R}^d} w(x + \delta y) \left[\frac{B(x + \delta y) - B(x)}{\delta} \cdot \nabla \eta(y) \right] \, dy. \end{aligned} \quad (78)$$

We denote by $\nabla \cdot B$ the Radon–Nikodym derivative of the measure $D \cdot B$ with respect to $\mathcal{L}^d$, that is $D \cdot B = D^s \cdot B + \nabla \cdot B \mathcal{L}^d$. Thus, we have $T_\delta = (r_\delta - w * \eta_\delta \nabla \cdot B) \mathcal{L}^d - w * \eta_\delta D^s \cdot B$, and

$$|T_\delta| = |r_\delta - w * \eta_\delta \nabla \cdot B| \mathcal{L}^d + |w * \eta_\delta| |D^s \cdot B|. \quad (79)$$

Using Proposition 4.3 we write r_δ as $r_{1,\delta} + r_{2,\delta}$, where

$$\begin{aligned} r_{1,\delta}(x) &:= - \int_{\mathbb{R}^d} w(x + \delta y) B_{1,\delta}(y)(x) \cdot \nabla \eta(y) \, dy, \\ r_{2,\delta}(x) &:= - \int_{\mathbb{R}^d} w(x + \delta y) B_{2,\delta}(y)(x) \cdot \nabla \eta(y) \, dy. \end{aligned}$$

Let σ be the weak* limit of a subsequence of $|T_\delta|$, and let $\varphi \in C_c(A)$ be a nonnegative test function. Then we get

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi \, d\sigma &\leq \limsup_{\delta \downarrow 0} \left\{ \int_{\mathbb{R}^d} \varphi(x) |r_{1,\delta}(x) - w * \eta_\delta(x) \nabla \cdot B(x)| \, dx \right. \\ &\quad + \int_{\mathbb{R}^d} \varphi(x) |r_{2,\delta}(x)| \, dx \\ &\quad \left. + \int_{\mathbb{R}^d} \varphi(x) |w * \eta_\delta(x)| |D^s \cdot B|(x) \, dx \right\}. \end{aligned} \quad (80)$$

We now analyze the behavior of the three integrals above.

First integral. From Proposition 4.3(a) and (c), and from the strong L^1_{loc} convergence of $w * \eta_\delta$ to w , it follows that

$$\begin{aligned} \lim_{\delta \downarrow 0} \int_{\mathbb{R}^d} \varphi(x) |r_{1,\delta}(x) - w * \eta_\delta(x) \nabla \cdot B(x)| \, dx \\ = \int_{\mathbb{R}^d} \varphi(x) \left| - \int_{\mathbb{R}^d} w(x) [\nabla \eta(y) \cdot \nabla B(x) \cdot y] \, dy \right. \\ \left. - w(x) \nabla \cdot B(x) \right| \, dx. \end{aligned} \quad (81)$$

Let $\mathcal{B}_{ij}(x)$ be the components of $\nabla B(x)$. For every $x \in \mathbb{R}^d$ we then compute

$$\begin{aligned} \int_{\mathbb{R}^d} w(x) [\nabla \eta(y) \cdot \nabla B(x) \cdot y] dy &= w(x) \sum_{i,j} \mathcal{B}_{ij}(x) \int_{\mathbb{R}^d} \partial_{y_i} \eta(y) y_j dy \\ &= -w(x) \sum_i \mathcal{B}_{ii}(x) \int_{\mathbb{R}^d} \eta(y) dy \\ &= -w(x) \nabla \cdot B(x), \end{aligned}$$

and therefore (81) vanishes.

Second integral. From now on, δ is assumed to be so small that if $\text{supp } \varphi + \text{supp } \eta_\delta \subset A$. Let us write $D^s B = M|D^s B|$, set $K_t := \{\varphi \geq t\}$ and write

$$\int_{\mathbb{R}^d} \varphi(x) |r_{2,\delta}(x)| dx = \int_0^\infty \int_{K_t} |r_{2,\delta}(x)| dx dt. \quad (82)$$

Note that $K_t = \emptyset$ for $t > \|\varphi\|_{C^0} =: T$ and $K_t \subset \text{supp}(\varphi) =: \Gamma$ for $t > 0$. On the other hand $\int_\Gamma |r_{2,\delta}(x)| dx$ is bounded by a constant C independent of δ by Proposition 4.3(c). This means that the functions $t \mapsto \int_{K_t} |r_{2,\delta}(x)| dx$ are bounded by the L^1 function $t \mapsto C \mathbb{1}_{[0,T]}(t)$. Hence, by the dominated convergence theorem,

$$\limsup_{\delta \downarrow 0} \int_{\mathbb{R}^d} \varphi(x) |r_{2,\delta}(x)| dx \leq \int_0^\infty \left\{ \limsup_{\delta \downarrow 0} \int_{K_t} |r_{2,\delta}(x)| dx \right\} dt. \quad (83)$$

Next, fix any compact set K , and consider

$$\int_K |r_{2,\delta}(x)| dx \leq \|w\|_{L^\infty(A)} \int_{\text{supp}(\eta)} \int_K |B_{2,\delta}(y)(x) \cdot \nabla \eta(y)| dx dy. \quad (84)$$

By the bound (c) in Proposition 4.3, the function

$$y \mapsto \int_K |B_{2,\delta}(y)(x) \cdot \nabla \eta(y)| dx \quad (85)$$

is uniformly bounded for $y \in \text{supp}(\eta)$. Hence, again by the dominated convergence theorem,

$$\begin{aligned} \limsup_{\delta \downarrow 0} \int_K |r_{2,\delta}(x)| dx \\ \leq \|w\|_{L^\infty(A)} \int_{\mathbb{R}^d} \left\{ \limsup_{\delta \downarrow 0} \int_K |B_{2,\delta}(y)(x) \cdot \nabla \eta(y)| dx \right\} dy. \end{aligned} \quad (86)$$

For any fixed y , use Remark 4.4 to get $B_{2,\delta}(y)(x) \cdot \nabla \eta(y) = [B \cdot \nabla \eta(y)]_{2,\delta}(y)(x)$. By Proposition 4.3(b), we then conclude

$$\limsup_{\delta \downarrow 0} \int_K |r_{2,\delta}(x)| \, dx \leq \|w\|_{L^\infty(A)} \int_{\mathbb{R}^d} |D^s(B \cdot \nabla \eta(y)) \cdot y|(K) \, dy. \quad (87)$$

On the other hand,

$$|D^s(B \cdot \nabla \eta(y)) \cdot y|(K) = \int_K |\nabla \eta(y) \cdot M(x) \cdot y| \, d|D^s B|(x). \quad (88)$$

Using (86)–(88), and exchanging the order of integration, we get

$$\begin{aligned} \limsup_{\delta \downarrow 0} \int_K |r_{2,\delta}(x)| \, dx \\ \leq \|w\|_{L^\infty(A)} \int_K \left[\int_{\mathbb{R}^d} |\nabla \eta(y) \cdot M(x) \cdot y| \, dy \right] d|D^s B|(x). \end{aligned} \quad (89)$$

Plugging (89) into (83), and recalling the definition of $\Lambda(M, \eta)$, we get

$$\begin{aligned} \limsup_{\delta \downarrow 0} \int_{\mathbb{R}^d} \varphi(x) |r_{2,\delta}(x)| \, dx \\ \leq \|w\|_{L^\infty(A)} \int_0^\infty \int_{K_t} \varphi(x) \Lambda(M(x), \eta) \, d|D^s B|(x) \, dt \\ = \|w\|_{L^\infty(A)} \int \varphi(x) \Lambda(M(x), \eta) \, d|D^s B|(x). \end{aligned} \quad (90)$$

Third integral. Finally, we have

$$\begin{aligned} \lim_{\delta \downarrow 0} \int_{\mathbb{R}^d} \varphi(x) |w * \eta_\delta(x)| \, d|D^s \cdot B|(x) \\ \leq \|w\|_{L^\infty(A)} \int_{\mathbb{R}^d} \varphi(x) \, d|D^s \cdot B|(x). \end{aligned} \quad (91)$$

Conclusion. From (80), (81), (90) and (91), we get

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi \, d\sigma \leq \|w\|_{L^\infty(A)} \int_{\mathbb{R}^d} \varphi(x) \Lambda(M(x), \eta) \, d|D^s B|(x) \\ + \|w\|_{L^\infty(A)} \int_{\mathbb{R}^d} \varphi(x) \, d|D^s \cdot B|(x) \end{aligned} \quad (92)$$

for every nonnegative $\varphi \in C_c(A)$, which implies the desired estimate

$$\sigma \llcorner A \leq \|w\|_{L^\infty(A)} \Lambda(M, \eta) |D^s B| + \|w\|_{L^\infty(A)} |D^s \cdot B|. \quad \square$$

4.3. Bouchut's lemma and Alberti's lemma

The following lemma was first proved by Bouchut in [15] and it was the starting point of Ambrosio's original proof of his commutator estimate (see [2]).

LEMMA 4.7 (Bouchut). *Let*

$$K := \left\{ \eta \in C_c^\infty(B_1(0)) \text{ such that } \eta \geq 0 \text{ is even, and } \int_{B_1(0)} \eta = 1 \right\}. \quad (93)$$

If $D \subset K$ is dense with respect to the strong $W^{1,1}$ topology, then for every $\xi, \chi \in \mathbb{R}^d$ we have

$$\inf_{\eta \in D} \Lambda(\chi \otimes \xi, \eta) = |\langle \xi, \chi \rangle| = |\operatorname{tr}(\chi \otimes \xi)|. \quad (94)$$

However, Ambrosio's original proof made use of the difficult rank-one theorem. Recently, Alberti has proposed an elementary proof of the following generalization of Bouchut's lemma.

LEMMA 4.8 (Alberti). *Let K be as in Lemma 4.7 and let M be a $d \times d$ matrix. Then*

$$\inf_{\eta \in D} \Lambda(M, \eta) = |\operatorname{tr} M|. \quad (95)$$

PROOF OF LEMMA 4.7. Set $M := \chi \otimes \xi$. Note that, since the map $\eta \in C_c^\infty(B_1(0)) \mapsto \Lambda(M, \eta)$ is continuous with respect to the strong $W^{1,1}$ topology, it is sufficient to prove that

$$\inf_{\eta \in K} \Lambda(M, \eta) = |\operatorname{tr} M|, \quad (96)$$

where K is the set in (93).

If $d = 2$ we can fix an orthonormal basis of coordinates z_1, z_2 in such a way that $\xi = (a, b)$ and $\chi = (0, c)$. Consider the rectangle $R_\varepsilon := [-\varepsilon/2, \varepsilon/2] \times [-1/2, 1/2]$ and consider the kernel $\eta_\varepsilon := \frac{1}{\varepsilon} \mathbb{1}_{R_\varepsilon}$. Let $\zeta \in K$ and denote by ζ_δ the family of mollifiers generated by ζ . Clearly $\eta_\varepsilon * \zeta_\delta \in K$ for $\varepsilon + \delta$ small enough.

Denote by $\nu = (\nu_1, \nu_2)$ the unit normal to ∂R_ε and recall that

$$\lim_{\delta \downarrow 0} \left| \frac{\partial(\eta_\varepsilon * \zeta_\delta)}{\partial z_i} \right| \rightharpoonup^* \frac{|\nu_i|}{\varepsilon} \mathcal{H}^1 \llcorner \partial R_\varepsilon \quad (97)$$

in the sense of measures.

Thus we can compute

$$\begin{aligned} \limsup_{\delta \downarrow 0} \Lambda(M, \eta_\varepsilon * \zeta_\delta) &\leq \limsup_{\delta \downarrow 0} \int_{\mathbb{R}^2} (|az_1| + |bz_2|) |c| \left| \frac{\partial(\eta_\varepsilon * \zeta_\delta)}{\partial z_2} \right| dz_1 dz_2 \\ &= \frac{2|c|}{\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} \left(|az_1| + \frac{|b|}{2} \right) dz_1 = |ac| \frac{\varepsilon}{2} + |bc|. \end{aligned}$$

Note that $bc = \operatorname{tr} M$. Thus, if we define the convolution kernels $\lambda_{\varepsilon, \delta} := \eta_\varepsilon * \zeta_\delta$ we get

$$\limsup_{\varepsilon \downarrow 0} \limsup_{\delta \downarrow 0} \Lambda(M, \eta_\varepsilon * \zeta_\delta) \leq |\operatorname{tr} M|. \quad (98)$$

For $d \geq 2$ we consider a system of coordinates $x_1, x_2, \dots, x_d$ such that $\eta = (a, b, 0, \dots, 0)$, $\xi = (0, c, 0, \dots, 0)$ and we define the convolution kernels

$$\lambda_{\varepsilon, \delta}(x) := [\eta_\varepsilon * \zeta_\delta](x_1, x_2) \cdot \zeta(x_3) \cdot \dots \cdot \zeta(x_d).$$

Then (98) holds as well and we conclude that, for any d , we have

$$\inf_{\eta \in K} \Lambda(M, \eta) \leq |\operatorname{tr} M|.$$

On the other hand, for every $\eta \in K$ and every $d \times d$ matrix M , we have

$$\begin{aligned} \Lambda(M, \eta) &\geq \left| \int_{B_1(0)} \langle M \cdot y, \nabla \eta(y) \rangle dy \right| = \left| \sum_{k,j} M_{jk} \int_{B_1(0)} y_j \frac{\partial \eta}{\partial z_k}(y) dy \right| \\ &= \left| - \sum_{k,j} M_{jk} \int_{B_1(0)} \delta_{jk} \eta(y) dy \right| = |\operatorname{tr} M|. \end{aligned} \quad (99)$$

This concludes the proof. □

The proof of the second lemma follows mainly [3].

PROOF OF LEMMA 4.8. As in the first proof, we note that it is sufficient to prove that

$$\inf_{\eta \in K} \Lambda(M, \eta) = |\operatorname{tr} M|, \quad (100)$$

and that the lower bound $\inf_{\eta \in K} \Lambda(M, \eta) \geq |\operatorname{tr} M|$ follows immediately from (99) (the argument leading to (99) does need the assumption $M = \chi \otimes \xi$). Therefore it remains to show the upper bound. Again by the identity $\langle M \cdot z, \nabla \eta(z) \rangle = \operatorname{div}(M \cdot z \eta(z)) - \operatorname{tr} M \eta(z)$, it suffices to show that for every $T > 0$ there exists $\eta \in K$ such that

$$\int_{\mathbb{R}^n} |\operatorname{div}(M \cdot z \eta(z))| dz \leq \frac{2}{T}. \quad (101)$$

Given a smooth nonnegative convolution kernel θ with compact support, we claim that the function

$$\eta(z) = \frac{1}{T} \int_0^T \theta(e^{-tM} \cdot z) e^{-t \operatorname{tr} M} dt$$

has the required properties. Here e^{tM} is the matrix

$$\sum_{i=0}^{\infty} \frac{t^i M^i}{i!}.$$

Thus $e^{tM} \cdot z$ is just the solution of the ODE $\dot{\gamma} = M \cdot \gamma$ with initial condition $\gamma(0) = z$, and $e^{-t \operatorname{tr} M}$ is the determinant of e^{-tM} . The usual change of variables yields

$$\begin{aligned} \int \eta(z) \varphi(z) dz &= \frac{1}{T} \int_0^T \int \varphi(z) \theta(e^{-tM} \cdot z) e^{-t \operatorname{tr} M} dz dt \\ &= \frac{1}{T} \int_0^T \int \varphi(e^{tM} \cdot \zeta) \theta(\zeta) d\zeta dt \end{aligned} \quad (102)$$

for any integrable bounded φ . Hence $\eta \mathcal{L}^d$ is the time average of the pushforward of the measure $\theta \mathcal{L}^d$ along the trajectories of $\dot{\gamma} = M \cdot \gamma$. This is the point of view taken in [3] to prove (101), for which we argue with the direct computations shown before.

Note that

$$\operatorname{div}(M \cdot z \eta(z)) = \frac{1}{T} \int_0^T \operatorname{div}(M \cdot z \theta(e^{-tM} \cdot z)) e^{-t \operatorname{tr} M} dt.$$

We compute

$$\begin{aligned} &\operatorname{div}(M \cdot z \theta(e^{-tM} \cdot z)) e^{-t \operatorname{tr} M} \\ &= \operatorname{tr} M \theta(e^{-tM} \cdot z) e^{-t \operatorname{tr} M} + \langle M \cdot z, e^{-tM} \cdot \nabla \theta(e^{-tM} \cdot z) \rangle e^{-t \operatorname{tr} M} \\ &= -\frac{d}{dt} (e^{-t \operatorname{tr} M}) \theta(e^{-tM} \cdot z) + \langle e^{-tM} \cdot M \cdot z, \nabla \theta(e^{-tM} \cdot z) \rangle e^{-t \operatorname{tr} M} \\ &= -\frac{d}{dt} (e^{-t \operatorname{tr} M}) \theta(e^{-tM} \cdot z) - \left\langle \frac{d}{dt} (e^{-tM} \cdot z), \nabla \theta(e^{-tM} \cdot z) \right\rangle e^{-t \operatorname{tr} M} \\ &= -\frac{d}{dt} (e^{-t \operatorname{tr} M}) \theta(e^{-tM} \cdot z) - \frac{d}{dt} (\theta(e^{-tM} \cdot z)) e^{-t \operatorname{tr} M} \\ &= -\frac{d}{dt} (\theta(e^{-tM} \cdot z) e^{-t \operatorname{tr} M}). \end{aligned}$$

Thus

$$\begin{aligned}
 \int_{\mathbb{R}^d} |\operatorname{div}(M \cdot z \eta(z))| \, dz &= \int_{\mathbb{R}^d} \frac{1}{T} \left| \int_0^T \operatorname{div}(M \cdot z \theta(e^{-tM} \cdot z)) e^{-t \operatorname{tr} M} \, dt \right| \, dz \\
 &= \int_{\mathbb{R}^d} \frac{1}{T} \left| \int_0^T \frac{d}{dt} (\theta(e^{-tM} \cdot z) e^{-t \operatorname{tr} M}) \, dt \right| \, dz \\
 &= \int_{\mathbb{R}^d} \frac{1}{T} |\theta(e^{-TM} \cdot z) e^{-T \operatorname{tr} M} - \theta(z)| \, dz \\
 &\leq \frac{1}{T} \left(\int_{\mathbb{R}^d} \theta(e^{-TM} \cdot z) e^{-T \operatorname{tr} M} \, dz + \int_{\mathbb{R}^d} \theta(z) \, dz \right) \\
 &= \frac{1}{T} \left(\int_{\mathbb{R}^d} \theta(\zeta) \, d\zeta + \int_{\mathbb{R}^d} \theta(z) \, dz \right) = \frac{2}{T},
 \end{aligned}$$

where in the last line we changed variables as in (102). This shows (101) and concludes the proof. $\square$

4.4. Proof of Theorem 4.1

We finally come to the proof of Theorem 4.1.

PROOF OF THEOREM 4.1. Let η be any smooth even convolution kernel. Set $\sigma_\delta := |T_\delta^i|$. From Proposition 4.6 we know that the total variation of these measures is uniformly bounded. Thus, recalling the computation of Section 4, and in particular (66), we conclude that $D \cdot (H(w)B)$ is a measure. Next, set

$$\alpha := D \cdot (H(w)B) - \left(H(w) - \sum_{i=1}^d \frac{\partial H}{\partial v_i}(w) w^i \right) D^a \cdot B$$

and let σ be the weak* limit of any subsequence of the measures $\{\sigma_\delta\}$. Then, from (68) we get

$$|\alpha| \leq C\sigma + C|D^s \cdot B|. \quad (103)$$

According to Proposition 4.6(b), this gives $|\alpha| \ll |D^s B|$, and thus we have $|\alpha| = g|D^s B|$ for some nonnegative Borel function g . Denote by M the Radon–Nikodym derivative of $D^s B$ with respect to $|D^s B|$. Then $|D^s \cdot B| = \operatorname{tr} M |D^s B|$. Thus, from (68) and (76) we conclude

$$g(x) \leq C(|\operatorname{tr} M(x)| + A(M(x), \eta)) \quad \text{for } |D^s B| \text{-a.e. } x. \quad (104)$$

Note that (104) holds for any even convolution kernel η . Let K be as in Lemma 4.8 and choose a countable set $D \subset K$ which is dense in the $W^{1,1}$ topology. Then

$$g(x) \leq C \left(|\operatorname{tr} M(x)| + \inf_{\eta \in D} \Lambda(M(x), \eta) \right) \quad \text{for } |D^s B| \text{-a.e. } x. \quad (105)$$

Therefore, from Lemma 4.8 we conclude

$$g(x) \leq C |\operatorname{tr} M(x)|,$$

which implies $|\alpha| \leq C |D^s \cdot B|$. Following the argument, one can readily check that C depends only on $R := \|w\|_\infty$ and $\|H\|_{C^1(B_R(0))}$. $\square$

REMARK 4.9. In this last step, the original proof of Ambrosio in [2] used Bouchut's lemma and Alberti's rank-one theorem (Theorem 2.13). Indeed, by Theorem 2.13 there exist two Borel vector-valued maps χ, ξ such that $M(x) = \chi(x) \otimes \xi(x)$ for $|D^s B|$ -a.e. x . Therefore, using this information one might rewrite (104) and (105) as

$$g(x) \leq C \left(|\operatorname{tr} M(x)| + \Lambda(\chi(x) \otimes \xi(x), \eta) \right) \quad \text{for } |D^s B| \text{-a.e. } x \quad (106)$$

and

$$g(x) \leq C \left(|\operatorname{tr} M(x)| + \inf_{\eta \in D} \Lambda(\chi(x) \otimes \xi(x), \eta) \right) \quad \text{for } |D^s B| \text{-a.e. } x. \quad (107)$$

From (107) it suffices to apply Lemma 4.7 to get

$$g(x) \leq C |\operatorname{tr} M(x)|.$$

5. Existence, uniqueness and stability for the Keyfitz and Kranzer system

In this section we consider the Cauchy problem for the Keyfitz and Kranzer system

$$\begin{cases} \partial_t u^i + \sum_{\alpha=1}^m \partial_{x_\alpha} (g^\alpha(|u|) u^i) = 0, \\ u^i(0, \cdot) = \bar{u}^i(\cdot). \end{cases} \quad (108)$$

Before stating the main theorem, we recall the notion of entropy solution of a scalar conservation law and the classical theorem of Kruzhkov, which provides existence, stability and uniqueness of entropy solutions to the Cauchy problem for scalar laws.

DEFINITION 5.1. Let $g \in W_{\text{loc}}^{1,\infty}(\mathbb{R}, \mathbb{R}^m)$. A pair (h, q) of functions $h \in W_{\text{loc}}^{1,\infty}(\mathbb{R}, \mathbb{R})$, $q \in W_{\text{loc}}^{1,\infty}(\mathbb{R}, \mathbb{R}^m)$ is called an *entropy-entropy flux pair* relative to g if

$$q' = h' g' \quad \mathcal{L}^1\text{-almost everywhere on } \mathbb{R}. \quad (109)$$

If, in addition, h is a convex function, then we say that (h, q) is a *convex entropy–entropy flux pair*. A weak solution $\rho \in L^\infty(\mathbb{R}_t^+ \times \mathbb{R}_x^m)$ of

$$\begin{cases} \partial_t \rho + D_x \cdot [g(\rho)] = 0, \\ \rho(0, \cdot) = \bar{\rho}(\cdot) \end{cases} \quad (110)$$

is called an *entropy solution* if $\partial_t[h(\rho)] + D_x \cdot [q(\rho)] \leq 0$ in the sense of distributions for every convex entropy–entropy flux pair (h, q) .

In what follows, we say that $\rho \in L^\infty(\mathbb{R}^+ \times \mathbb{R}^m)$ has a *strong trace* $\bar{\rho}$ at 0 if for every bounded $\Omega \subset \mathbb{R}^n$ we have

$$\lim_{T \downarrow 0} \frac{1}{T} \int_{[0, T] \times \Omega} |\rho(t, x) - \bar{\rho}(x)| \, dx \, dt = 0.$$

THEOREM 5.2 (Kruzhkov [36]). *Let $g \in W_{\text{loc}}^{1, \infty}(\mathbb{R}, \mathbb{R}^m)$ and $\bar{\rho} \in L^\infty$. Then there exists a unique entropy solution ρ of (110) with a strong trace at $t = 0$. If in addition $\bar{\rho} \in BV_{\text{loc}}(\mathbb{R}^m)$, then, for every open set $A \subset \subset \mathbb{R}^m$ and for every $T \in]0, \infty[$, there exists an open set $A' \subset \subset \mathbb{R}^m$ (whose diameter depends only on A , T , g and $\|\bar{\rho}\|_\infty$) such that*

$$\|\rho\|_{BV([0, T] \times A)} \leq \|\bar{\rho}\|_{BV(A')}. \quad (111)$$

Often, in what follows we will use the terminology *Kruzhkov solution* for entropy solutions of (110) with a strong trace at $t = 0$.

REMARK 5.3. In many cases the requirement that ρ has strong trace at 0 is not necessary. Indeed, when g is sufficiently regular and satisfies suitable assumptions of genuine nonlinearity, Vasseur proved in [39] that any entropy solution has a strong trace at 0.

We are now ready to introduce the particular class of weak solutions of (108) for which we are able to prove existence, uniqueness, and continuous dependence with respect to the initial data.

DEFINITION 5.4. A weak solution u of (108) is called a *renormalized entropy solution* if $|u|$ is an Kruzhkov solution of the scalar law

$$\begin{cases} \partial_t \rho + \sum_{\alpha=1}^m \partial_{x_\alpha} (g^\alpha(\rho) \rho) = 0, \\ \rho(0, \cdot) = \bar{\rho}(\cdot). \end{cases} \quad (112)$$

In the class of renormalized entropy solutions we have the following well-posedness theorem for bounded initial data $\bar{u}$ such that $|\bar{u}| \in BV_{\text{loc}}$.

THEOREM 5.5. *Let $g \in W_{\text{loc}}^{1, \infty}(\mathbb{R}, \mathbb{R}^k)$ and $|\bar{v}| \in L^\infty \cap BV_{\text{loc}}$. Then there exists a unique renormalized entropy solution u of (108). If $\bar{v}^j$ is a sequence of initial data such that*

- (a) $|\bar{v}^j| \leq C$ for some constant C ,
 - (b) for every bounded open set Ω , there is a constant $C(\Omega)$ such that $\|\bar{v}^j\|_{BV(\Omega)} \leq C(\Omega)$,
 - (c) $\bar{v}^j \rightarrow \bar{v}$ strongly in L^1_{loc} ,
- then the corresponding renormalized entropy solutions converge strongly in L^1_{loc} to u .

The suggestion of using the terminology *renormalized entropy solutions* has been taken from Frid [32]. This terminology is more appropriate than the one of *entropy solutions* used in [8], because the usual notion of *entropy* (or *admissible*) solution of a hyperbolic system of conservation laws does not coincide with the one of renormalized entropy solutions. Let us recall the usual notion of entropy solution for systems (cf. Section 4.3 of [23]).

DEFINITION 5.6. Let $F^\alpha : \mathbb{R}^k \rightarrow \mathbb{R}^k$, $\alpha = 1, \dots, n$, be Lipschitz and consider the system

$$\partial_t u + \sum_{\alpha=1}^m \partial_{x_\alpha} [F^\alpha(u)] = 0, \quad u : \Omega \subset \mathbb{R}^+ \times \mathbb{R}^m \rightarrow \mathbb{R}^k. \quad (113)$$

A pair (H, Q) of functions $H \in W^{1,\infty}_{\text{loc}}(\mathbb{R}^k, \mathbb{R})$, $Q \in W^{1,\infty}_{\text{loc}}(\mathbb{R}^k, \mathbb{R}^m)$ is called a *convex entropy-entropy flux pair* for the system (113) if H is convex and if $DQ^\alpha = DH \cdot DF^\alpha$, for every $\alpha \in \{1, \dots, m\}$.

A distributional solution u of (113) supplemented by the initial condition

$$u(0, \cdot) = \bar{u}(\cdot)$$

is called an *entropy solution* if for every convex entropy-entropy flux pair (H, Q) and for every smooth test function $\psi \geq 0$,

$$\begin{aligned} & \int_{t>0} \int_{\mathbb{R}^m} [\partial_t \psi(t, z) H(u(t, z)) + \nabla_z \psi(t, z) \cdot Q(u(t, z))] dt dz \\ & + \int_{\mathbb{R}^m} \psi(0, z) \eta(\bar{u}(z)) dz \geq 0. \end{aligned} \quad (114)$$

The (nonpositive) entropy production measure

$$\partial_t [H(u)] + D_x \cdot [Q(u)]$$

will be denoted by μ_H .

The system of Keyfitz and Kranzer corresponds to the particular case $F(u) = u \otimes g(|u|)$. We will later show that, under suitable assumptions on g , for every convex entropy H for (108) there exists a convex function $h : \mathbb{R}^k \rightarrow \mathbb{R}$ and a Lipschitz function $\widehat{H} : \mathbf{S}^{k-1} \rightarrow \mathbb{R}$ such that

$$H(v) = h(|v|) + |v| \widehat{H}\left(\frac{v}{|v|}\right) \quad \text{for every } v \neq 0$$

(see Lemma 5.11 and compare with Lemma 1.1 of [32]).

Using this lemma we will show that if u is a renormalized entropy solution, then u is an entropy solution in the sense of Definition 5.6.

PROPOSITION 5.7. *Assume $g \in C^1$ and $\mathcal{L}^1(\{s > 0: g'(s) = 0\}) = 0$. Then every renormalized entropy solution of (108) is an entropy solution.*

Actually we expect this statement to be true even if we drop the assumption $\mathcal{L}^1(\{s > 0: g'(s) = 0\}) = 0$. However Lemma 5.11 does not hold in general and therefore a more refined approach is required.

Clearly, another natural question is whether the opposite inclusion

$$\{\text{entropy solutions}\} \subset \{\text{renormalized entropy solutions}\}$$

holds. It can be shown that, already in one space dimension, there exist entropy solutions of (108) which are not renormalized entropy solutions (see for instance [23]). This is essentially caused by the degeneration at the origin of the hyperbolicity of the Keyfitz and Kranzer system. However under appropriate assumptions on the initial data, it is reasonable to expect that any entropy solution coincides with the unique renormalized entropy solution. In particular we propose the following conjecture.

CONJECTURE 5.8. Let u be a bounded entropy solution of (108) and denote by C the closure of the convex hull of its essential image. If $0 \notin C$ or if it is an extremal point of C , then u is a renormalized entropy solution.

A partial answer to this conjecture is given by the following proposition.

PROPOSITION 5.9. *Let $f \in W_{\text{loc}}^{1,\infty}$ and $\bar{u} \in L^\infty(\mathbb{R}^m, \mathbb{R}^k)$. Denote by C be the closure of the convex hull of the essential image of u and assume that*

- (a) *either $0 \notin C$ or it is an extremal point of C ;*
- (b) *u is a bounded entropy solution of (108);*
- (c) *$u \in BV([0, T] \times \Omega)$ for some $T > 0$ and for some bounded open $\Omega \subset \mathbb{R}^m$.*

Then u is a renormalized entropy solution of (108) on $]0, T[\times \Omega$.

5.1. Proof of Theorem 5.5

The proof of Theorem 5.5 follows from the theory of transport equations for nearly incompressible fields via Ambrosio's renormalization theorem. More precisely, the key point is the following lemma.

LEMMA 5.10. *Let $\rho \in L^\infty([0, \infty[\times \mathbb{R}^m)$, $b \in L^\infty([0, \infty[\times \mathbb{R}^m, \mathbb{R}^m)$ be such that*

- *$b, \rho \in BV([0, T] \times K)$ for every compact set K ;*
- *(29) holds, that is, $\partial_t \rho + D_x \cdot (\rho b) = 0$;*
- *$\rho(0, \cdot) \in BV_{\text{loc}}$.*

Then the pair (b, ρ) has the renormalization property.

PROOF. Recall that, from the trace properties of BV functions we have

$$\lim_{T \downarrow 0} \frac{1}{T} \int_0^T \int_K |\rho(t, x) - \rho(0, x)| + |b(t, x) - b(0, x)| \, dx \, dt = 0$$

for every compact set $K \subset \mathbb{R}^m$. We define $\hat{\rho} \in BV_{\text{loc}}(\mathbb{R}^{m+1})$, $\hat{b} \in BV_{\text{loc}}(\mathbb{R}^{m+1})$ by setting

$$\hat{\rho}(t, x) = \begin{cases} \rho(0, x) & \text{if } t \leq 0, \\ \rho(t, x) & \text{if } t > 0, \end{cases} \quad \text{and} \quad \hat{b}(t, x) = \begin{cases} 0 & \text{if } t \leq 0, \\ b(t, x) & \text{if } t > 0. \end{cases}$$

Now, let $u \in L^\infty([0, \infty[\times \mathbb{R}^m)$ and $\bar{u} \in L^\infty(\mathbb{R}^m)$ be such that

$$\begin{cases} \partial_t(\rho u) + D_x \cdot (b \rho u) = 0, \\ [\rho u](0, \cdot) = \rho(0, \cdot) \bar{u}, \end{cases} \quad (115)$$

and define

$$\hat{u}(t, x) = \begin{cases} \bar{u}(x) & \text{if } t < 0, \\ u(t, x) & \text{if } t \geq 0. \end{cases}$$

Then $\partial_t(\hat{\rho} \hat{u}) + D_x \cdot (\hat{\rho} \hat{u} \hat{b}) = 0$ distributionally on $\mathbb{R}^{m+1}$. Thus, if we apply Theorem 4.1 to $B = (\hat{\rho}, \hat{\rho} \hat{b})$, $\hat{u}$ and $H(v) = v^2$, since $D \cdot B = 0$, we conclude that

$$\partial_t(\hat{u}^2 \hat{\rho}) + D_x \cdot (\hat{u}^2 \hat{\rho} \hat{b}) = 0.$$

From Lemma 3.7 we have that, up to change $\hat{\rho} \hat{u}^2$ on a set of measure zero, the map $t \mapsto \hat{\rho}(t, \cdot) \hat{u}^2(t, \cdot)$ is weakly continuous. Since for $t < 0$ we have $\hat{\rho}(t, \cdot) \hat{u}^2(t, \cdot) = \rho(0, \cdot) \bar{u}^2(\cdot)$ and for $t > 0$ we have $\hat{\rho}(t, \cdot) \hat{u}^2(t, \cdot) = \rho(t, \cdot) u^2(t, \cdot)$, we conclude that $\rho(0, \cdot) \bar{u}^2(\cdot)$ is the trace at $t = 0$ of the function ρu^2 . Thus we get

$$\begin{cases} \partial_t(\rho u^2) + D_x \cdot (b \rho u^2) = 0, \\ [\rho u^2](0, \cdot) = \rho(0, \cdot) \bar{u}^2. \end{cases}$$

With an analogous argument one shows that if

$$\begin{cases} \partial_t(\rho u) + D_x \cdot (b \rho u) = 0, \\ [\rho u](0, \cdot) = \rho(0, \cdot) \bar{u}, \\ [\rho u](T, \cdot) = \rho(T, \cdot) \hat{u}, \end{cases} \quad (116)$$

then $v = u^2$ solves

$$\begin{cases} \partial_t(\rho v) + D_x \cdot (b \rho v) = 0, \\ [\rho v](0, \cdot) = \rho(0, \cdot) \bar{u}^2, \\ [\rho v](T, \cdot) = \rho(T, \cdot) \hat{u}^2. \end{cases}$$

□

PROOF OF THEOREM 5.5.

Existence. Let g and $\bar{u}$ be as in the statement. First of all, let ρ be the Kruzhkov solution of

$$\begin{cases} \partial_t \rho + D_x \cdot (\rho g(\rho)) = 0, \\ \rho(0, \cdot) = |\bar{u}|(\cdot). \end{cases} \quad (117)$$

Then, Kruzhkov's theory gives $\|\rho\|_\infty \leq \|\bar{u}\|_\infty$ and $\rho \in BV([0, T] \times K)$ for every compact set. Since g is locally Lipschitz, $g(\rho) \in BV([0, T] \times K)$. Therefore, by Lemma 5.10, the pair $(b, \rho) := (g(\rho), \rho)$ has the renormalization property.

Next let $\bar{\theta} \in L^\infty(\mathbb{R}^n, \mathbf{S}^{k-1})$ be any function such that $\bar{u} = |\bar{u}|\bar{\theta}$ and apply Proposition 3.13 to get a bounded solution θ of

$$\begin{cases} \partial_t(\rho\theta) + D_x \cdot (\theta \otimes (\rho g(\rho))) = 0, \\ [\rho\theta](0, \cdot) = \bar{\rho}(0, \cdot)\bar{\theta}(\cdot). \end{cases} \quad (118)$$

Consider the continuous function $H: \mathbb{R}^k \rightarrow [0, \infty[$ given by $H(v) := |v|$. Applying Lemma 5.10 and Proposition 3.10 we conclude that

$$\begin{cases} \partial_t(\rho|\theta|) + D_x \cdot (\rho|\theta|g(\rho)) = 0, \\ [\rho|\theta|](0, \cdot) = \bar{\rho}(0, \cdot)|\bar{\theta}(\cdot)| = \bar{\rho}(0, \cdot). \end{cases}$$

Thus, from Proposition 3.13, it follows $\rho|\theta| = \rho$. Therefore, if we define $u := \rho\theta$, we have $|u| = \rho$ and hence

- $|u|$ is a Kruzhkov solution of (117);
- u solves

$$\begin{cases} \partial_t u + D_x \cdot (u \otimes g(|u|)) = 0, \\ u(0, \cdot) = \bar{u}. \end{cases}$$

Uniqueness. The uniqueness follows easily from the uniqueness of Kruzhkov solutions for the Cauchy problem of scalar conservation laws and from Proposition 3.13.

Stability. The stability follows directly from the stability of Kruzhkov solutions for scalar conservation laws and from Corollary 3.21. $\square$

5.2. Renormalized entropy solutions are entropy solutions

In this subsection we prove Proposition 5.7. The key remark is the following lemma (see [32]).

LEMMA 5.11. Assume $g \in C^1([0, \infty[, \mathbb{R}^k)$ and $\mathcal{L}^1(\{s > 0: g'(s) = 0\}) = 0$. Consider the map $F^\alpha \in W_{\text{loc}}^{1,\infty}(\mathbb{R}^k, \mathbb{R}^k)$ given by $F^\alpha(u) = g^\alpha(|u|)u$. If (H, Q) is a convex entropy–entropy flux pair in the sense of Definition 5.6, then there exist a convex $h \in W_{\text{loc}}^{1,\infty}([0, \infty[)$ and an $\widehat{H} \in W^{1,\infty}(\mathbb{S}^{k-1})$ such that

$$H(u) = h(|u|) + |u|\widehat{H}\left(\frac{u}{|u|}\right) \quad \text{for any } u \neq 0.$$

In order to simplify the notation, in what follows, if $\widehat{H}: \mathbb{S}^{k-1} \rightarrow \mathbb{R}$ is a bounded function, we extend the function

$$\mathbb{R}^k \setminus \{0\} \ni u \rightarrow |u|\widehat{H}\left(\frac{u}{|u|}\right) \in \mathbb{R}$$

by defining as 0 its value at 0. Clearly this extension is Lipschitz whenever $\widehat{H}$ is Lipschitz.

REMARK 5.12. Note that at least the assumption that $\{g' = 0\}$ has empty interior is needed in order to conclude Lemma 5.11. Indeed, assume $]a, b[\subset \{g' = 0\}$. Then g is constantly equal to some vector γ on that interval. Consider any convex function $H \in C^2(\mathbb{R}^k)$ with the following properties

- $H = 0$ on $\{0 \leq |v| \leq (a+b)/2\}$,
- $H(v) = |v|$ on $\{v \in \mathbb{R}^k: |v| \geq b\}$,

and let Q be given by

- $Q(v) = H(v)\gamma$ for $0 \leq |v| \leq b$;
- $Q(v) = |v|f(|v|)$ for $|v| \geq b$.

Then (H, Q) is a convex entropy–entropy flux pair, but H is not necessarily of the form $h(|u|) + |u|\widehat{H}(u/|u|)$.

Nonetheless we expect that the conclusion of Proposition 5.7 holds in general. Indeed, if $g' = 0$ on $[a, b]$ and u is a solution of (108) such that $a \leq |u| \leq b$, then u solves k decoupled transport equations with constant coefficients. Thus u is trivially an entropy solution. However, a more refined analysis would be needed if the range of $|u|$ contains both intervals where g' vanishes and intervals where $g' \neq 0$.

Lemma 5.11 easily implies Proposition 5.7.

PROOF OF PROPOSITION 5.7. Let g be as in the proposition, let u be any renormalized entropy solution and let H, Q be an entropy–entropy flux pair. We apply Lemma 5.11 to get $H(u) = h(|u|) + |u|\widehat{H}(u/|u|)$, where h is convex and $\widehat{H}$ is Lipschitz. Let $q \in W^{1,\infty}(\mathbb{R})$ be such that $q(0) = Q(0)$ and $q'(r) = h'(r)g'(r)r + h'(r)g(r)$. Then it follows easily that $Q(u) = q(|u|) + |u|g(|u|)\widehat{H}(u/|u|)$. Let $\psi \in C_c^\infty(-\infty, \infty \times \mathbb{R}^m)$ be any test function. Since $|u|$ is a Kruzkov solution of

$$\begin{cases} \partial_t \rho + D_x \cdot (g(\rho)\rho) = 0, \\ \rho(0, \cdot) = |u(0, \cdot)|, \end{cases}$$

we have

$$\begin{aligned} & \int_{t>0} \int_{\mathbb{R}^m} [\partial_t \psi(t, z) h(|u(t, z)|) + \nabla_z \psi(t, z) \cdot q(|u(t, z)|)] dt dz \\ & + \int_{\mathbb{R}^m} \psi(0, z) h(|\bar{u}(z)|) dz \geq 0. \end{aligned} \quad (119)$$

Moreover, from the renormalization property applied to θ we must have

$$\begin{aligned} & \int_{t>0} \int_{\mathbb{R}^m} |u|(t, x) \widehat{H}\left(\frac{u(t, x)}{|u|(t, x)}\right) \\ & \times [\partial_t \psi(t, z) + \nabla_z \psi(t, z) \cdot g(|u(t, z)|)] dt dz \end{aligned} \quad (120)$$

$$+ \int_{\mathbb{R}^m} \psi(0, z) |\bar{u}(z)| \widehat{H}\left(\frac{\bar{u}(z)}{|\bar{u}(z)|}\right) dz = 0. \quad (121)$$

Summing (119) and (120) we conclude (114). This completes the proof. $\square$

PROOF OF LEMMA 5.11. If g , H , and Q satisfy the assumptions of the lemma, then Q is a Lipschitz function and the identity

$$\nabla Q^\alpha(v) = \nabla H(v) \cdot \nabla(g^\alpha(|v|)|v|) \quad (122)$$

is valid for $\mathcal{L}^k$ -a.e. $v \in \mathbb{R}^k \setminus \{0\}$.

Now consider a smooth system of coordinates $\omega_1, \dots, \omega_{k-1}$ on $\mathbf{S}^{k-1}$ and let $\omega_1, \dots, \omega_{k-1}, r$ be polar coordinates on $\mathbb{R}^k \setminus \{0\}$. It is not difficult to see that (122) becomes

$$\begin{cases} \partial_{\omega_i} Q^\alpha(r, \omega) = g^\alpha(r) \partial_{\omega_i} H(r, \omega), \\ \partial_r Q^\alpha(r, \omega) = ((g^\alpha)'(r)r + g^\alpha(r)) \partial_r H(r, \omega). \end{cases} \quad (123)$$

(In other words, $\omega_1, \dots, \omega_{k-1}, r$ is a coordinate system of Riemann invariants for the Keyfitz and Kranzer system; see [23] or [38] for the definition.)

These identities hold pointwise a.e. and hence (since Q and H are Lipschitz) in the sense of distributions. Therefore, from $\partial_{r\omega_i}^2 Q^\alpha = \partial_{\omega_i r}^2 Q^\alpha$ we conclude

$$\partial_r(g^\alpha(r) \partial_{\omega_i} H(r, \omega)) = \partial_{\omega_i} \{((g^\alpha)'(r)r + g^\alpha(r)) \partial_r H(r, \omega)\}. \quad (124)$$

Recall that H is convex, and hence its second derivatives are measures. Thus

$$\partial_r(g^\alpha(r) \partial_{\omega_i} H(r, \omega)) = (g^\alpha)'(r) \partial_{\omega_i} H(r, \omega) + g^\alpha(r) \partial_{r\omega_i}^2 H, \quad (125)$$

where the product $g^\alpha(r) \partial_{r\omega_i}^2 H$ makes sense because $g^\alpha(r)$ is continuous.

For the same reason, since $\partial_{r\omega_i}^2 H$ is a measure and $(g^\alpha)'(r)$ is continuous, a standard smoothing argument justifies

$$\partial_{\omega_i} \{ ((g^\alpha)'(r)r + g^\alpha(r)) \partial_r H(r, \omega) \} = (g^\alpha(r) + (g^\alpha)'(r)r) \partial_{r\omega_i}^2 H. \quad (126)$$

Comparing (124) with (125) and (126), we get

$$(g^\alpha)'(r) \partial_{\omega_i} H(r, \omega) + g^\alpha(r) \partial_{r\omega_i}^2 H = (g^\alpha(r) + (g^\alpha)'(r)r) \partial_{r\omega_i}^2 H$$

and hence

$$(g^\alpha)'(r) (r \partial_{r\omega_i}^2 H - \partial_{\omega_i} H) = 0. \quad (127)$$

If we set $p(r) := \sum_\alpha |(g^\alpha)'(r)|$, we obtain

$$p(r) (r \partial_{r\omega_i}^2 H - \partial_{\omega_i} H) = 0. \quad (128)$$

We claim that, since $\mathcal{L}^1(\{r: p(r) = 0\}) = 0$, we have

$$r \partial_{r\omega_i}^2 H - \partial_{\omega_i} H = 0 \quad \text{distributionally on } \mathbb{R}^k \setminus \{0\}. \quad (129)$$

Indeed, consider the measures $\mu := r \partial_{r\omega_i}^2 H$ and $\alpha := \mu - \partial_{\omega_i} H$ and let $\Omega \subset \mathbb{R}^2 \setminus \{0\}$ be the open set $\{x \in \mathbb{R}^k \setminus \{0\}: |p(|x|)| = 0\}$. Then $\alpha \equiv 0$ on Ω . Hence it suffices to show $|\alpha|(\mathbb{R}^2 \setminus \Omega) = 0$. Since $\mathcal{L}^k(\mathbb{R}^k \setminus (\{0\} \cup \Omega)) = 0$ and $\partial_{\omega_i} H \ll \mathcal{L}^k$, it suffices to show

$$|\mu|(\mathbb{R}^2 \setminus (\{0\} \cup \Omega)) = 0.$$

In order to prove this identity, recall that $\mu = r \partial_{\omega_i}(\partial_r H)$ and that $\partial_r H$ is a BV function, because H is convex. Consider for every $\tau > 0$ the function $\sigma_\tau(\omega) := \partial_r H(\tau, \omega)$. From the slicing theory of BV functions, it follows that $\sigma_\tau \in BV(\mathbf{S}^{k-1})$ for $\mathcal{L}^1$ -a.e. $\tau > 0$ and that

$$|\mu| = r \int_0^\infty |\partial_{\omega_i} \sigma_\tau| d\tau.$$

Thus, since $\mathcal{L}^1(\{\tau: p(\tau) = 0\}) = 0$, we have $|\mu|(\mathbb{R}^k \setminus (\{0\} \cup \Omega)) = 0$, which concludes the proof of (129).

Note that (129) can be rewritten as

$$r^2 \partial_r \left(\frac{\partial_{\omega_i} H}{r} \right) = 0$$

and hence we get that

$$\partial_{\omega_i} H(r, \omega) = r \psi_i(\omega)$$

for some locally bounded function ψ_i . Let N be the north pole of $\mathbf{S}^{k-1}$, i.e. the point corresponding to $(1, 0, \dots, 0)$ for some orthonormal system of coordinates on $\mathbb{R}^k \supset \mathbf{S}^{k-1}$. Consider the restriction $H|_{\mathbf{S}^{k-1}}$ of H on $\mathbf{S}^{k-1}$ and let $\widehat{H} \in C(\mathbf{S}^{k-1})$ be given by $\widehat{H}(\omega) = H|_{\mathbf{S}^{k-1}}(\omega) - H(N)$. Then $\partial_{\omega_i}(r\widehat{H}(\omega)) = r\psi_i(\omega)$. Therefore

$$\partial_{\omega_i}(H(r, \omega) - r\widehat{H}(\omega)) = 0$$

and hence $H(r, \omega) - r\widehat{H}(\omega) = h(r)$ for some function h . Moreover, we have

$$h(r) = H(r, N) - r\widehat{H}(N) = H(r, N).$$

That is, h is given by the restriction of H to the half-line $\{(\tau, 0, \dots, 0) : \tau \geq 0\}$. Therefore h is necessarily convex. $\square$

5.3. Proof of Proposition 5.9

Let u and Ω be as in the statement. Define $\rho := |u|$ and $\bar{\rho} := |\bar{u}|$. The goal is to show that ρ is an entropy solution of the scalar law

$$\begin{cases} \partial_t \rho + D_x \cdot [g(\rho)\rho] = 0, \\ \rho(0, \cdot) = |\bar{u}|, \end{cases} \quad (130)$$

in $]0, T[\times \Omega$.

Actually it is sufficient to show that ρ is a *weak* solution of (130) in $]0, T[\times \Omega$. Indeed, note that for every $h : \mathbb{R}^+ \rightarrow \mathbb{R}$ which is convex and increasing, $h(|u|)$ is a convex entropy for the system (108) (the entropy flux is of the form $q(|u|)$ for q such that $q' = h'g'$). Thus we have

$$\begin{aligned} & \int_{t>0} \int_{\mathbb{R}^m} [\partial_t \psi(t, z) h(\rho(t, z)) + \nabla_x \psi(t, z) \cdot q(\rho(t, z))] dt dz \\ & + \int_{\mathbb{R}^m} \psi(0, z) h(\bar{\rho}(z)) dz \geq 0 \end{aligned} \quad (131)$$

for every nonnegative smooth test function ψ . Moreover, if ρ is a weak solution of (130) in $]0, T[\times \Omega$, L a linear function $L : \mathbb{R} \rightarrow \mathbb{R}$ and $Q : \mathbb{R} \rightarrow \mathbb{R}^m$ the map given by $Q = (L(g^1), \dots, L(g^m))$, then

$$\begin{aligned} & \int_{t>0} \int_{\mathbb{R}^m} [\partial_t \psi(t, z) L(\rho(t, z)) + \nabla_x \psi(t, z) \cdot Q(\rho(t, z))] dt dz \\ & + \int_{\mathbb{R}^m} \psi(0, z) L(\bar{\rho}(z)) dz = 0, \end{aligned} \quad (132)$$

for every test function $\psi \in C_c^\infty([0, T] \times \Omega)$. Given any convex function ξ we can write it as $L + h$, where L is an appropriate linear function and h is increasing on the half-line

$\mathbb{R}^+$. Thus, summing (131) and (132), we conclude that ρ satisfies the entropy inequality for ξ and for every nonnegative $\psi \in C_c^\infty([-T, T] \times \Omega)$, and hence that ρ is an entropy solution of (130) in $]0, T[\times \Omega$.

We now come to the proof that ρ is a weak solution of (130), which we split in several steps.

Step 1. Recall that ρ is a weak solution of (130) in $]0, T[\times \Omega$ if it satisfies the identity

$$\begin{aligned} \int_{t>0} \int_{\mathbb{R}^m} \rho(t, z) [\partial_t \psi(t, z) + g(\rho(t, z)) \cdot \nabla_x \psi(t, z)] dz dt \\ + \int_{\mathbb{R}^m} \psi(0, z) \bar{\rho}(z) dz = 0 \end{aligned} \quad (133)$$

for every $\psi \in C_c^\infty([-T, T] \times \Omega)$.

Recall that $\|u\|_{BV(\Omega \times]0, T])}$ is finite. Hence, we claim that thanks to the trace properties of BV functions, in order to prove (133) it suffices to check that

$$\text{the Radon measure } \mu = \partial_t \rho + D_x \cdot (\rho g(\rho)) \text{ vanishes on }]0, T[\times \Omega. \quad (134)$$

Indeed, by a standard approximation argument we get the following estimate for every $t < T$:

$$\begin{aligned} \int_0^t \int_{\Omega} |u(\tau, z) - \bar{u}(z)| dz d\tau &\leq \int_0^t |\partial_t u|([0, \tau[\times \Omega) d\tau \\ &\leq t |\partial_t u|([0, t[\times \Omega). \end{aligned}$$

From this we conclude

$$\int_0^t \int_{\Omega} |\rho(\tau, z) - \bar{\rho}(z)| dz d\tau \leq t |\partial_t u|([0, t[\times \Omega). \quad (135)$$

Fix $\psi \in C_c^\infty([-T, T] \times \Omega)$ and let $\{\chi_i\} \subset C^\infty([0, T])$ be such that

- $\chi_i = 1$ for $t \geq 2/i$;
- $\chi_i = 0$ for $t \leq 1/i$;
- $0 \leq \chi_i' \leq 4i$.

Then, $\psi \chi_i$ is compactly supported in $]0, T[\times \Omega$ and from (134) we get

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^m} \chi_i(\tau) \rho(\tau, z) [\partial_t \psi(\tau, z) + g(\rho(\tau, z)) \cdot \nabla_x \psi(\tau, z)] dz d\tau \\ + \int_0^{2/k} \int_{\mathbb{R}^m} \chi_i'(\tau) \rho(\tau, z) \psi(\tau, z) dz d\tau = 0. \end{aligned} \quad (136)$$

As $i \uparrow \infty$, the first integral in (136) converges to

$$\int_0^T \int_{\mathbb{R}^m} \rho(\tau, z) [\partial_t \psi(\tau, z) + g(\rho(\tau, z)) \cdot \nabla_x \psi(\tau, z)] dz d\tau.$$

Concerning the second integral, we recall that $\int_0^{2/i} \chi'_i = 1$ and we write:

$$\begin{aligned}
 & \left| \int_0^{2/i} \int_{\mathbb{R}^m} \chi'_i(\tau) \rho(\tau, z) \psi(\tau, z) \, dz \, d\tau - \int_{\mathbb{R}^m} \bar{\rho}(z) \psi(0, z) \, dz \right| \\
 &= \left| \int_0^{2/i} \int_{\mathbb{R}^m} \chi'_i(\tau) [\rho(\tau, z) \psi(\tau, z) - \bar{\rho}(z) \psi(0, z)] \, dz \, d\tau \right| \\
 &\leq 4i \int_0^{2/i} \int_{\mathbb{R}^m} |\rho(\tau, z) \psi(\tau, z) - \bar{\rho}(z) \psi(0, z)| \, d\tau \, dz \\
 &\leq 4i \|\rho\|_\infty \int_0^{2/i} \int_{\mathbb{R}^m} |\psi(\tau, z) - \psi(0, z)| \, d\tau \, dz \\
 &\quad + 4i \|\psi\|_\infty \int_0^{2/i} \int_{\mathbb{R}^m} |\rho(\tau, z) - \rho(0, z)| \, d\tau \, dz.
 \end{aligned}$$

Note that, for $i \uparrow \infty$, the first term tends to 0 because ψ is smooth. Thanks to (135) the second term is bounded by

$$C |\partial_t u| \left(\left[0, \frac{2}{i} \right] \times \Omega \right), \quad (137)$$

where C is a constant independent of t , and Ω is a bounded set. Since $|\partial_t u|$ is Radon measure, we conclude that the expression (137) tends to 0 for $i \uparrow \infty$. Thus we conclude that

$$\lim_{i \uparrow \infty} \int_0^{2/i} \int_{\mathbb{R}^m} \chi'_i(\tau) \rho(\tau, z) \psi(\tau, z) \, dz \, d\tau = \int_{\mathbb{R}^m} \bar{\rho}(z) \psi(0, z) \, dz.$$

Hence, passing into the limit in (136) we get (133). Therefore, we are left with the task of proving (134).

Step 2 We wish to use the entropy inequalities and to apply Theorem 2.11 to conclude that μ is supported on the jump set (or shock set) J_u . However this is not possible since the function $|u|$ is not C^1 in the origin (compare with Remark 2.12). We approximate this function uniformly with smooth C^1 convex functions of the form $h_n(|u|)$. Clearly, also these functions are entropies for the system of Keyfitz and Kranzer and their entropy fluxes are of the form $q_n(|u|)$ for some functions $q_n(t)$ which converge uniformly to $tf(t)$.

Let $v: J_u \rightarrow \mathbb{R}^m$ be a Borel vector field and $\zeta: J_u \rightarrow \mathbb{R}$ be a nonnegative Borel function such that $(\zeta, v)/\sqrt{\zeta^2 + |v|^2}$ is normal to J_u $\mathcal{H}^m$ -a.e. Then, the chain rule of Volpert gives that

$$\begin{aligned}
 & \partial_t [h_n(\rho)] + D_x \cdot [q_n(\rho)] \\
 &= (\zeta^2 + |v|^2)^{-1/2} \\
 &\quad \times [(h_n(|u^+|) - h_n(|u^-|))\zeta + (q_n(|u^+|) - q_n(|u^-|)) \cdot v] \mathcal{H}^m \llcorner J_u.
 \end{aligned}$$

Passing to the limit in n we get:

$$\begin{aligned} \mu &= (\zeta^2 + |v|^2)^{-1/2} \\ &\times [(|u^+| - |u^-|)\zeta \\ &+ (|u^+|g(|u^+|) - |u^-|g(|u^-|)) \cdot v] \mathcal{H}^m \llcorner J_u. \end{aligned} \quad (138)$$

Thus, we must prove that

$$(\zeta + g(|u^+|) \cdot v)|u^+| = (\zeta + g(|u^-|) \cdot v)|u^-| \quad \mathcal{H}^m\text{-a.e. on } J_u. \quad (139)$$

In what follows, for the sake of simplicity, we will drop the $\mathcal{H}^m$ -a.e.Ó.

Since u is a weak solution of (108), when $F(v) := g(|v|) \otimes v$ is C^1 we can apply Theorem 2.11 to get

$$(g(|u^+|) \cdot v + \zeta)u^+ = (g(|u^-|) \cdot v + \zeta)u^-. \quad (140)$$

In order to derive (140) when 0 is a singularity for DF we approximate F with $F_n := g(h_n(u)) \otimes u$. Then we get

$$\begin{aligned} \partial_t u + D_x \cdot (F_n(u)) &= D^d u + DF_n(\tilde{u}) \cdot D^d u \\ &+ [(u^+ - u^-)\zeta + (F(u^+) - F(u^-)) \cdot v] \mathcal{H}^m \llcorner J_u. \end{aligned} \quad (141)$$

Clearly, the left-hand side converges to $0 = \partial_t u + D_x \cdot (F(u))$. Moreover, the second term of the right-hand side converges to

$$[(g(|u^+|) \cdot v + \zeta)u^+ - (g(|u^-|) \cdot v + \zeta)u^-] \mathcal{H}^m \llcorner J_u$$

in the sense of measures.

Note that the approximations F_n can be chosen in such a way that DF_n are locally uniformly bounded. In this case, let σ be any weak* limit of any subsequence of $DF_n(\tilde{u}) \cdot D^d u$. Since $|DF_n \cdot D^d u| \leq C|D^d u|$, this weak* limit satisfies $\sigma \ll |D^d u|$. On the other hand, passing into the limit in (141) we get

$$0 = \sigma + [(g(|u^+|) \cdot v + \zeta)u^+ - (g(|u^-|) \cdot v + \zeta)u^-] \mathcal{H}^m \llcorner J_u.$$

Since $|D^d u|(J_u) = 0$, we conclude that (140) holds $\mathcal{H}^m$ -a.e. on J_u .

From (140) we get

$$|g(|u^+|) \cdot v + \zeta||u^+| = |g(|u^-|) \cdot v + \zeta||u^-|. \quad (142)$$

If $|u^+|$ (or $|u^-|$) vanishes, (139) follows trivially. Hence, after setting $\rho^\pm := |u^\pm|$ we restrict our attention to the subset of J_u given by $G := \{\rho^+ \neq 0 \neq \rho^-\}$. On this set we define $\theta^\pm := u^\pm / \rho^\pm$ and we note that (140) becomes

$$[(g(\rho^+) \cdot v + \zeta)]\rho^+\theta^+ = [(g(\rho^-) \cdot v + \zeta)]\rho^-\theta^-. \quad (143)$$

Since $\theta^\pm \in \mathbf{S}^{k-1}$ we conclude that, either $\theta^+ = \theta^-$ or $\theta^+ = -\theta^-$. In the next step we will prove that, if D is the closure of the convex hull of the essential image of $u|_{[0,T] \times \mathbb{R}^m}$, then either $0 \notin D$ or 0 is an extremal point of D . This rules out the alternative $\theta^+ = -\theta^-$. Therefore we conclude that $\theta^+ = \theta^-$ on G , from which (139) easily follows.

Step 3. In order to complete the proof it remains to show that, if D denotes the closure of the convex hull of the essential image of $u|_{[0,T] \times \mathbb{R}^m}$, then either the origin is not contained in D , or it is an extremal point of D . Recalling (a), this property is true for the closure C of the convex hull of the essential image of $\bar{u}$. Choose $\xi_1, \dots, \xi_k$ unit vectors of $\mathbb{R}^k$ such that

$$C \subset \{x \mid x \cdot \xi_i \leq 0 \text{ for every } i\}$$

and 0 is an extremal point of $\{x \mid x \cdot \xi_i \leq 0 \text{ for every } i\}$. We will show that the essential image of u is contained in $\{x \mid x \cdot \xi_i \leq 0\}$ for every i .

Fix i and denote by $H: \mathbb{R}^k \rightarrow \mathbb{R}$, $Q: \mathbb{R}^k \rightarrow \mathbb{R}^m$ the functions

$$H(v) := \begin{cases} 0 & \text{if } \xi_i \cdot v \leq 0, \\ \xi_i \cdot v & \text{otherwise,} \end{cases} \quad Q(v) := f(|v|)H(v).$$

Note that (H, Q) is a convex entropy flux pair. Clearly $H(\bar{u}) = 0$ and thus the boundary term in the entropy inequality (114) disappears. Thus, if we set $w := H(u)$ and $b := Q(u)$ we get that

$$\begin{cases} \partial_t w + D_x \cdot b \leq 0, \\ w(0, \cdot) = 0. \end{cases}$$

Note that there exists a constant C such that $|b| \leq Cw$. Therefore we can apply Lemma 3.17 to conclude $w \equiv 0$. This completes the proof.

6. Blow-up of the BV norm for the Keyfitz and Kranzer system

In one space dimension, the fundamental result of Glimm (see [23]) gives the existence of BV entropy solutions for (108) if one starts with initial data which have sufficiently small total variation. Moreover, from Proposition 5.9 we get that, when the convex hull of the essential image of the initial data $\bar{u}$ does not contain the origin (or the origin is an extremal point of it), such solution is the unique renormalized entropy solution.

Hence it is natural to ask whether renormalized entropy solutions u of (108) enjoy BV regularity when the whole initial datum $\bar{u}$ (and not only its modulus) belongs to BV. In analogy with the one-dimensional case, one could ask if such regularity holds at least for small times and when $\bar{u}$ is close to a constant different from 0, in both the L^∞ and the BV norms. We will show that this is not the case. More precisely, we will show the following.

THEOREM 6.1. *Let $k \geq 2$, $m \geq 3$, $g \in C_{\text{loc}}^3$ and let $c \in \mathbb{R}^k \setminus \{0\}$ such that $g'(|c|) \neq 0$. Then there exists a sequence of initial data $\bar{u}_n: \mathbb{R}^m \rightarrow \mathbb{R}^k$ such that*

- $\|\bar{u}_n - c\|_{BV(\mathbb{R}^m)} + \|\bar{u}_n - c\|_\infty \rightarrow 0$ for $n \uparrow \infty$;
- $\bar{u}_n = c$ on $\mathbb{R}^m \setminus B_R(0)$ for some $R > 0$ independent of n ;
- if u_n is any bounded entropy solution of (108) with initial data $\bar{u}_n$, then there exists $r > 0$ (independent of n) such that $\|u_n\|_{BV([0, T] \times B_r(0))} = \infty$ for every positive T .

When $m = 2$ the same statement holds if in addition we assume that $g''(|c|)$ is parallel to $g'(|c|)$ (or vanishes).

We remark that the system of KeyPtz and Kranzer, in contrast to general hyperbolic systems of conservation laws, has remarkably many features. Indeed consider the system of conservation laws

$$\partial_t u + D_x \cdot [F(u)] = 0 \quad u : \Omega \subset \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^k, \quad (144)$$

where $F : \mathbb{R}^k \rightarrow \mathbb{R}^{k \times m}$ is a C^1 function. In what follows we will use the notation $F = (F^1, \dots, F^m)$, where each F^i is a map from $\mathbb{R}^k$ to $\mathbb{R}^k$. The KeyPtz and Kranzer system corresponds to the choice $F(v) = v \otimes g(|v|)$, where $g \in C^1(\mathbb{R}, \mathbb{R}^m)$. (Note that in this case the requirement $F \in C^1$ implies $g'(0) = 0$. However, in the rest of the forthcoming sections we will not impose this condition, since it is not needed in any of the proofs.) Therefore the KeyPtz and Kranzer system falls into the category of symmetric systems of conservation laws, i.e., the systems (144) for which $DF^i(v)$ is a symmetric matrix for every i and for every $v \in \mathbb{R}^k$.

It is known, by a result of Rauch based on a previous paper of Brenner for linear hyperbolic systems (see [16] and [37]), that certain type of BV estimates (and L^p estimates for $p \neq 2$) fail for all the systems (144) which do not satisfy the commutator conditions

$$DF^i(v) \cdot DF^j(v) = DF^j(v) \cdot DF^i(v) \quad \text{for every } v \in \mathbb{R}^k. \quad (145)$$

When $m = 2$, it was proved in [22] that (145) is also sufficient to get L^p estimates for every $p \leq 2$ and, under additional conditions, also for $p = \infty$.

Note that the KeyPtz and Kranzer system does satisfy Rauch's commutator condition (145). Moreover, we remark that when (145) does not hold, Rauch's result implies that estimates of a certain kind are not available, but it does not exclude BV regularity.

6.1. Preliminary lemmas

In this section we collect some facts which will be used in the proof of Theorem 6.1.

Riemann problem for scalar laws. Let us consider the Cauchy problem

$$\begin{cases} \partial_t \rho + D_x \cdot [h(\rho)] = 0, \\ \rho(0, \cdot) = \bar{\rho}, \end{cases} \quad \rho : \mathbb{R}^+ \times \mathbb{R}^m \rightarrow \mathbb{R}, \quad (146)$$

where $h: \mathbb{R} \rightarrow \mathbb{R}^m$ is of class C^3 . Fix $\beta, \gamma, \alpha \in \mathbb{R}$, set $\varepsilon := \max\{|\alpha - \beta|, |\alpha - \gamma|\}$, and choose

$$\bar{\rho}(x_1, \dots, x_m) = \begin{cases} \beta & \text{for } x_m < 0, \\ \gamma & \text{for } x_m > 0. \end{cases}$$

Consider the entropy solution ρ of (146). It is easy to see that ρ depends only on t and x_m . For each $T > 0$ define:

$$\xi := \max\{x_m \mid \rho(T, \cdot, x_m) = \beta\}, \quad (147)$$

$$\zeta := \min\{x_m \mid \rho(T, \cdot, x_m) = \gamma\}. \quad (148)$$

Then the following lemma has an elementary proof.

LEMMA 6.2. *Let $T > 0$ and $\alpha \in \mathbb{R}$ be given. For any real α and β , set ε, ξ and ζ as before. If we denote by $(h^m)'$ and $(h^m)''$ the m -th components of the vector-valued functions h' and h'' , then there exist constants C and δ (depending only on h) such that*

$$\begin{aligned} & \max\{|\xi - T(h^m)'(\alpha)|, |\zeta - T(h^m)'(\alpha)|\} \\ & \leq 2|(h^m)''(\alpha)|\varepsilon + C\varepsilon^2 \quad \text{for } \varepsilon \leq \delta. \end{aligned} \quad (149)$$

Regular Lagrangian flows. Let u be a renormalized entropy solution of (108). Assume that the initial data $\bar{u}$ is bounded away from the origin, i.e., that $|\bar{u}| \geq c > 0$. Then, from the maximum principle for scalar conservation laws, it turns out that the renormalized entropy solution u is bounded away from zero as well, i.e., that $|u| \geq c > 0$. Hence the angular parts $\bar{\theta} := \bar{u}/|\bar{u}|$, $\theta := u/|u|$ are well defined and solve the transport equation (118).

Let Φ be the unique regular Lagrangian flow given by Theorem 3.22:

$$\begin{cases} \frac{d}{dt} \Phi(s, x) = g(\rho(s, \Phi(s, x))), \\ \Phi(0, x) = x. \end{cases} \quad (150)$$

Then the following holds.

PROPOSITION 6.3. *There exists a locally bounded map $\Psi: \mathbb{R}^+ \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ such that $\Phi(s, \Psi(s, x)) = \Psi(s, \Phi(s, x)) = x$ for $\mathcal{L}^{m+1}$ -a.e. (s, x) . Moreover, $\theta(t, x) = \bar{\theta}(\Psi(t, x))$.*

PROOF. Let $\{f_n\} \subset C^\infty$ be a uniformly bounded sequence such that $f_n \rightarrow g(\rho)$ in L^1_{loc} and $\{\rho_n\} \subset C^\infty$ a sequence of positive functions such that

- $\|\rho_n^{-1}\|_\infty + \|\rho_n\|_\infty$ is uniformly bounded;
- $\rho_n \rightarrow \rho$ and $\rho_n(0, \cdot) \rightarrow \rho(0, \cdot)$ in L^1_{loc} ;
- $\partial_t \rho_n + D_x \cdot (\rho_n f_n) = 0$.

These approximating sequences can be constructed as in the proof of the existence part of Theorem 3.22 (in particular see Step 1). Let Φ_n be the solutions of the ODEs

$$\begin{cases} \frac{d}{dt} \Phi_n(s, x) = f_n(s, \Phi_n(s, x)), \\ \Phi_n(0, x) = x. \end{cases} \quad (151)$$

Then for some constant C we have $C^{-1} \leq \det \nabla_x \Phi_n \leq C$. Thus, if we let $\Psi_n : \mathbb{R}^+ \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ be such that $\Psi(t, \Phi(t, x)) = (t, x)$, then $\{\|\Psi_n\|_{L^\infty([0, T] \times K)}\}$ for every $T > 0$ and every compact set $K \subset \mathbb{R}^m$.

From Theorem 3.22, Φ_n converges to Φ strongly in L^1_{loc} . Moreover, from the proof of the stability property of Theorem 3.22, it follows easily that $\Psi_n \rightarrow \Psi$ strongly in L^1_{loc} to some bounded map Ψ . From these convergence and from the bounds

$$C^{-1} \leq \det \nabla_x \Phi_n \leq C \quad C^{-1} \leq \det \nabla_x \Psi_n \leq C,$$

it is easy to conclude that $\Psi(t, \Phi(t, x)) = \Phi(t, \Psi(t, x)) = x$ for $\mathcal{L}^{m+1}$ -a.e. (t, x) .

Set $\tilde{\theta}(t, x) := \tilde{\theta}(\Psi(t, x))$, then, for $\mathcal{L}^m$ -a.e. x , the function $\tilde{\theta}(\cdot, \Phi(\cdot, x))$ is constant. Therefore, by Proposition 3.5, we get that $\tilde{\theta}$ solves (118). From Corollary 3.14 we conclude that $\tilde{\theta} = \theta$. $\square$

PROPOSITION 6.4. *For $\mathcal{L}^m$ -a.e. x we have that:*

- (a) $\Phi(\cdot, x)$ is Lipschitz (and hence it is differentiable in t for $\mathcal{L}^1$ -a.e. t);
- (b) $(t, \Phi(t, x))$ is a point of approximate continuity of ρ for $\mathcal{L}^1$ -a.e. t ;
- (c) $\frac{d}{dt} \Phi(t, x) = g(\rho(t, \Phi(t, x)))$ for $\mathcal{L}^1$ -a.e. t .

PROOF.

Step 1. Consider again two sequences of smooth maps $\{f_n\}, \{\rho_n\}$ as in the proof of the previous proposition. Denote by Φ_n the solutions of (151) and set $J_n := \det(\nabla_x \Phi_n)$. From Liouville's theorem it follows that $\partial_t J_n + \text{div}(f_n J_n) = 0$. Since $J_n(0, \cdot) = 1$, the maximum principle of Proposition 3.13 applied to the continuity equation $\partial_t w + \text{div}(f_n w) = 0$ yields that $C^{-1} \rho_n \leq J_n \leq C \rho_n$, and hence $C^{-2} \leq J_n \leq C^2$.

Recall that $\Phi_n \rightarrow \Phi$ strongly in L^1_{loc} . Since for every x the curves $\Phi_n(\cdot, x)$ are uniformly Lipschitz, we conclude that $\Phi(\cdot, x)$ is a Lipschitz curve for $\mathcal{L}^m$ -a.e. x . This gives (a).

Step 2. Next, fix a t and a subsequence (not relabeled) of $\Phi_n(t, \cdot)$ which converges to $\Phi(t, \cdot)$ in $L^1_{\text{loc}}(\mathbb{R}^m)$ (such a subsequence exists for $\mathcal{L}^1$ -a.e. t). Let $E \subset \mathbb{R}^m$ be an open set. It is not difficult to show that

$$\mathcal{L}^m(\Phi(t, \cdot)^{-1}(E)) \leq \limsup_{n \uparrow \infty} \mathcal{L}^m(\Phi_n(t, \cdot)^{-1}(E)) \leq C^2 \mathcal{L}^m(E). \quad (152)$$

Hence, for $\mathcal{L}^1$ -a.e. t , this bound holds for every open set E . This property gives that for $\mathcal{L}^1$ -a.e. t , $\Phi(t, \cdot)^{-1}$ maps sets of measure zero into sets of measure zero. Thus (b) follows from the fact that ρ is almost everywhere approximately continuous.

Step 3. The strong convergence of Φ_n implies that, if $h_n \in C(\mathbb{R} \times \mathbb{R}^m)$ converges locally uniformly to $h \in C(\mathbb{R} \times \mathbb{R}^m)$, then $h_n(\cdot, \Phi_n)$ converges to $h(\cdot, \Phi)$ strongly in L^1_{loc} . If $h_n \rightarrow h$ strongly in L^1_{loc} and it is uniformly bounded, applying Egorov's theorem we find a closed set E such that h_n converges locally uniformly to h on E and $\mathcal{L}^{m+1}(\mathbb{R} \times \mathbb{R}^m \setminus E)$ is as small as desired. Recall that Φ_n is locally uniformly bounded. From Step 2 it follows that $h_n(\cdot, \Phi_n)$ converges strongly to $h(\cdot, \Phi)$.

Step 4. Since Φ_n solves (151) we have

$$\Phi_n(t, x) = x + \int_0^t f_n(\tau, \Phi_n(\tau, x)) \, d\tau. \quad (153)$$

Applying Step 3 to $h_n = f_n$ and $h = g(\rho)$ we get a subsequence (not relabeled) of $\{\Phi_n\}$ such that $f_n(\cdot, \Phi_n)$ converges to $g(\rho(\cdot, \Phi))$ pointwise a.e. on $\mathbb{R} \times \mathbb{R}^m$. From the dominated convergence theorem we get

$$\Phi(t, x) = x + \int_0^t g(\rho(\tau, \Phi(\tau, x))) \, d\tau \quad \text{for } \mathcal{L}^{m+1}\text{-a.e. } (t, x).$$

From this identity we easily conclude (c). □

6.2. Proof of Theorem 6.1

Theorem 6.1 is a corollary of Proposition 5.9 and of the following proposition.

PROPOSITION 6.5. *Let $k \geq 2$, $m \geq 3$ and $g \in C_{\text{loc}}^3$. Then, for every $c \in \mathbb{R}^k \setminus \{0\}$ such that $g'(|c|) \neq 0$, there exists a sequence of initial data $\bar{u}_n : \mathbb{R}^m \rightarrow \mathbb{R}^k$ such that*

- $\|\bar{u}_n - c\|_{BV(\mathbb{R}^m)} + \|\bar{u}_n - c\|_\infty \rightarrow 0$ for $n \uparrow \infty$;
- $\bar{u}_n = c$ on $\mathbb{R}^m \setminus B_R(0)$ for some $R > 0$ independent of n ;
- if u_n denotes the unique renormalized entropy solution of (108) with $u_n(0, \cdot) = \bar{u}_n$, then there exists $r > 0$ such that $u_n(t, \cdot) \notin BV(B_r(0))$ for every n and for every $t \in]0, 1[$.

When $m = 2$ the same statement holds if in addition $g''(|c|)$ is parallel to $g'(|c|)$ or $g''(|c|) = 0$.

PROOF OF THEOREM 6.1. Let $\bar{u}_n$ be the initial data of Proposition 6.5 and let $r > 0$ be such that the corresponding renormalized entropy solutions $u_n(t, \cdot)$ are not in $BV(B_r(0))$ for any $t \in]0, 1[$. Let $\hat{u}_n$ be any other entropy solution of (108) with the same initial data. For any $c > \|\bar{u}_n\|_\infty$, we apply the argument of Step 3 of the proof of Proposition 5.9 to the entropy $h(|u|) := (|u| - c)\mathbb{1}_{|u| \geq c}$. It turns out that $h(|u|) = 0$, from which we conclude $\|\hat{u}_n\|_\infty \leq \|\bar{u}_n\|_\infty$. Hence $\hat{u}_n$ is uniformly bounded.

Fix $T \in]0, 1[$ and let $\gamma \geq 0$ be the supremum of the nonnegative ROs such that $\hat{u}_n \in BV(]0, T[\times B_\gamma(0))$. We want to bound γ with a constant times r . From Proposition 5.9 we get that $\hat{u}_n$ is a renormalized entropy solution on $]0, T[\times B_\gamma(0)$. Therefore $\hat{\rho}_n := |\hat{u}_n|$ is a Kruzkov solution of

$$\begin{cases} \partial_t \hat{\rho}_n + D_x \cdot (\hat{\rho}_n g(\hat{\rho}_n)) = 0 & \text{on }]0, T[\times B_\gamma(0), \\ \hat{\rho}_n(0, \cdot) = \bar{\rho}_n. \end{cases}$$

From the finite speed of propagation of scalar conservation laws, it follows that there exists positive constants T_1 and γ_1 such that $\rho_n = \hat{\rho}_n$ on $]0, T_1[\times B_{\gamma_1}(0)$. Moreover, we can choose

$$\gamma_1 \geq c\gamma, \quad T_1 \geq cT, \quad (154)$$

where the constant $c > 0$ depends only on $\|\bar{u}_n\|_\infty$ on g .

Set $\hat{\theta}_n = \hat{u}_n/\hat{\rho}_n$ and $\theta_n = u_n/\rho_n$, with the convention that $\hat{\theta}_n = 0$ where $\hat{\rho}_n = 0$ and $\theta_n = 0$ where $\rho_n = 0$. Then $\hat{\theta}_n$ and θ_n solve both the transport equation

$$\begin{cases} \partial_t(\rho_n \omega) + D_x \cdot (\rho_n g(\rho_n) \omega) = 0 & \text{in }]0, T_1[\times B_{\gamma_1}(0), \\ [\rho_n \omega](0, \cdot) = \bar{u}_n. \end{cases}$$

Thus, by the renormalization property, we get that $w = |\theta_n - \hat{\theta}_n|$ solves

$$\begin{cases} \partial_t(\rho_n w) + D_x \cdot (\rho_n g(\rho_n) w) = 0 & \text{in }]0, T_1[\times B_{\gamma_1}(0), \\ [\rho_n w](0, \cdot) = 0. \end{cases}$$

From Lemma 3.17, we conclude that there exists two positive constants $\gamma_2 < \gamma_1$ and $T_2 < T_1$ such that $w = 0$ on $]0, T_2[\times B_{\gamma_2}(0)$, and that we can choose

$$\gamma_2 \geq c'\gamma_1, \quad T_2 \geq c'T_1, \quad (155)$$

where c' depends only on $\|\rho_n\|_\infty \leq \|\bar{u}_n\|_\infty$ and g .

Since $\|\bar{u}_n\|_\infty$ is uniformly bounded, the constants c and c' in (154) and (155) can be chosen independently of n . Recall that $u_n \notin BV([0, T_2] \times B_r(0))$. This implies the desired bound $\gamma < cc'r$. Indeed, if such a bound did not hold, then we would have $\gamma_2 \geq r$ and hence $u_n = \hat{u}_n$ on $]0, T_2[\times B_r(0)$. This would imply $u_n \in BV([0, T_2] \times B_r(0))$, which is a contradiction. $\square$

In the next section we will give a proof of Proposition 6.5. But first we consider the special case of system (108) when $g = (f, 0, \dots, 0)$, that is,

$$\begin{cases} \partial_t u + \partial_{x_1} [f(|u|)u] = 0, \\ u(0, \cdot) = u_0. \end{cases} \quad (156)$$

The following is a corollary of Proposition 6.5.

PROPOSITION 6.6. *Let $k \geq 2$, $m \geq 2$ and $c \in \mathbb{R}^k \setminus \{0\}$ be such that $f'(|c|) \neq 0$. Then there exists a sequence of initial data $\bar{u}_n : \mathbb{R}^m \rightarrow \mathbb{R}^k$ such that*

- $\|\bar{u}_n - c\|_{BV(\mathbb{R}^m)} + \|\bar{u}_n - c\|_\infty \rightarrow 0$ for $n \uparrow \infty$;
- $\bar{u}_n = c$ on $\mathbb{R}^m \setminus B_R(0)$ for some $R > 0$ independent of n ;
- if u_n denotes the unique renormalized entropy solution of (156) with $u_n(0, \cdot) = \bar{u}_n$, then there exists $r > 0$ such that $u_n(t, \cdot) \notin BV_{\text{loc}}(B_r(0))$ for every n and for every $t \in]0, 1[$.

Roughly speaking, the proof of Proposition 6.5 is based on the following remark: When $m = 3$ we can choose initial data, close to a constant, in such a way that the behavior of the renormalized entropy solutions of (108) is close to the behavior of solutions of (156). This seems to be no longer true for $m = 2$, unless $g''(|c|)$ is parallel to $g'(|c|)$ (or $g''(|c|) = 0$). Due to this remark, we choose to give a quick self-contained proof of Proposition 6.6.

REMARK 6.7. Concerning the behavior of u_n for large times, in the case of Proposition 6.6 one can construct initial data $\bar{u}_n$ such that $u_n(t, \cdot) \notin BV_{\text{loc}}$ for any positive time $t > 0$. In the case of Proposition 6.5 it is difficult to track what happens for large times, since in order to carry on our proof we need that the rarefaction waves generated by $|u_n|$ do not interact.

PROOF OF PROPOSITION 6.6. In the following, for any real number α , we denote by $[\alpha]$ the largest integer which is less than or equal to α .

For the sake of simplicity we prove the proposition when $m = 2$, $f'(|c|) = 1$, and $f(|c|) = 0$. Only minor adjustments are needed to handle the general case. To simplify the notation, on $\mathbb{R}^2$ we will use the coordinates (x, y) in place of (x_1, x_2) .

Let $\{m_i\}$ be a sequence of positive even numbers such that

$$\sum_i m_i 2^{-i} < \infty. \quad (157)$$

Let $\delta > 0$ be so small that:

- f is injective on $[|c| - 2\delta, |c| + 2\delta]$;
- $[-\delta, \delta] \subset f([|c| - 2\delta, |c| + 2\delta])$.

Then, for i sufficiently large, we define r_i as the unique number in $[-2\delta, 2\delta]$ such that $f(|c| + r_i) = 2^{-i}$. Notice that for i sufficiently large we have $r_i \leq 2^{-i+1}$. Set $\alpha = c/|c|$ and for every i choose an $\alpha_i \in \mathbf{S}^{k-1}$ such that $|\alpha_i - \alpha| = i^{-2}$.

Let I_i be the interval $[2^{-i}, 2^{-i+1}[$ and subdivide it in m_i equal subintervals

$$I_i^j := \left[2^{-i} + \frac{(j-1)2^{-i}}{m_i}, 2^{-i} + \frac{j2^{-i}}{m_i} \right], \quad j \in \{1, \dots, m_i\}.$$

Next define the functions $\psi_i : \mathbb{R}^2 \rightarrow \mathbf{S}^{k-1}$ as

$$\psi_i(x, y) := \begin{cases} \alpha_i & \text{if } y \in I_i \text{ and } [x2^i] \text{ is odd,} \\ \alpha & \text{otherwise,} \end{cases}$$

and the functions $\chi_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ as

$$\chi_i(x, y) := \begin{cases} r_i & \text{if } y \in I_i^j \text{ for } j \text{ even and } x \in [-M, M], \\ r_{i+1} & \text{if } y \in I_i^j \text{ for } j \text{ odd and } x \in [-M, M], \\ 0 & \text{otherwise.} \end{cases}$$

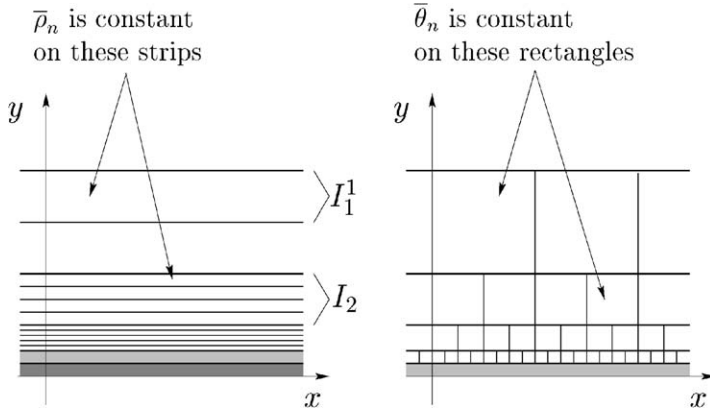


Fig. 1. Decomposition of the plane in open sets where $\bar{\rho}_n$ (resp. $\bar{\theta}_n$) is constant.

Here M is a positive real number which will be chosen later. Finally we define

$$\begin{aligned}\bar{\rho}_n &:= |c| + \sum_{i=n}^{\infty} \chi_i, \\ \bar{\theta}_n(x, y) &:= \begin{cases} \psi_i(x, y) & \text{if } y \in I_i \text{ for some } i \geq n \text{ and } x \in [-M, M], \\ \alpha & \text{otherwise,} \end{cases} \\ \bar{u}_n &:= \bar{\rho}_n \bar{\theta}_n.\end{aligned}$$

Figure 1 gives a picture of the partition of $\mathbb{R}^2$ on which we based the definition of $\bar{u}_n$.

Clearly $\|\bar{u}_n - c\|_{\infty} \leq |c| |\alpha_n - \alpha| + r_n$. Hence, as $n \uparrow \infty$ we have $\|\bar{u}_n - c\|_{\infty} \rightarrow 0$. Moreover, notice that $\bar{u}_n - c$ is supported on $[-M, M] \times [0, 1]$. From now on we assume that M will be chosen large than 1.

In order to show that

$$\|\bar{u}_n - c\|_{BV(\mathbb{R}^2)} \rightarrow 0$$

it is sufficient to show

$$\|\bar{\rho}_n - |c|\|_{BV([-2M, 2M]^2)} \rightarrow 0, \quad (158)$$

$$\|\bar{\theta}_n - \alpha\|_{BV([-2M, 2M]^2)} \rightarrow 0. \quad (159)$$

Note that

$$\begin{aligned}\|\bar{\rho}_n - |c|\|_{BV([-2M, 2M]^2)} &\leq 4\|\bar{u}_n - c\|_{\infty} M^2 + 2M \sum_{i \geq n} m_i r_i + (4M + 2)r_n, \\ &\leq 4\|\bar{u}_n - c\|_{\infty} M^2 + 4M \sum_{i \geq n} m_i 2^{-i} + (4M + 2)r_n,\end{aligned}$$

and since $\sum 2^{-i} m_i$ is summable, we get (158). Moreover,

$$\begin{aligned} \|\bar{\theta}_n - \alpha\|_{BV([-2M, 2M]^2)} &\leq 4\|\bar{\theta}_n - \alpha\|_{\infty} M^2 + 2M \sum_{i \geq n} 2^{-i} i^{-2} 2^i \\ &\quad + 2M \sum_{i \geq n} [i^{-2} + (i+1)^{-2}] + (4M+2)n^{-2} \end{aligned}$$

and the summability of $\sum i^{-2}$ gives (159).

Now we let u_n be the unique renormalized solution of (156). Recall that $\rho_n := |u_n|$ is the unique entropy solution of (117) with initial data $\bar{\rho}_n$, which in our case is given by

$$\begin{cases} \partial_t \rho_n + \partial_x (f(\rho_n) \rho_n) = 0, \\ \rho_n(0, \cdot) = \bar{\rho}_n. \end{cases}$$

Hence, if $\bar{\rho}_n$ did not depend on x , we would have $\rho_n(t, y, x) = \bar{\rho}_n(x, y)$. Since $\bar{\rho}_n$ is truncated, this is not true. However, $\bar{\rho}_n(\cdot, y)$ is constant on $[-M, M]$ and by the finite speed of propagation of scalar laws it follows that $\rho_n(t, x, y) = \bar{\rho}_n(x, y)$ if (t, x, y) belongs to the cone

$$\{\sqrt{y^2 + x^2} \leq c(M - t)\},$$

where c is a constant which depends only on $\|\bar{\rho}_n\|_{\infty}$. Thus, for every $\lambda > 1$, we can choose M large enough (but independent of n) so that

$$\rho_n(t, x, y) = \bar{\rho}_n(x, y) \quad \text{for } t \in [0, 1] \text{ and } (x, y) \in [-\lambda, \lambda] \times [0, 1].$$

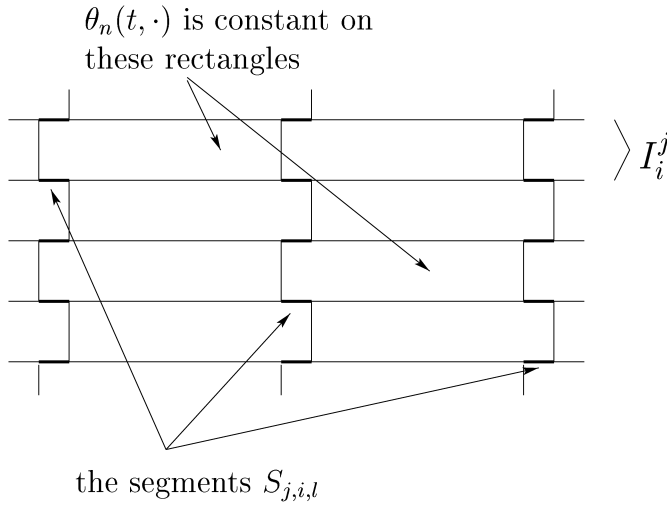
To find the angular part $\theta_n(t, x, y) := u_n/|u_n|(t, x, y)$ we use the fact that θ_n is constant on the curves $\Phi_n(\cdot, x)$, where Φ_n solves the ODEs

$$\begin{cases} \frac{d}{dt} \Phi_n(s, x, y) = g(\rho_n(s, \Phi_n(s, x, y))), \\ \Phi_n(0, x, y) = (x, y), \end{cases} \quad (160)$$

in the sense of Propositions 6.3 and 6.4. Hence it follows that, for $\mathcal{L}^3$ -a.e. (τ, x_1, y_1) there is $(x_0, y_0) \in \mathbb{R}^2$ such that:

- the curve $\Phi(\cdot, x_0, y_0)$ is Lipschitz;
- $\Phi(\tau, x_0, y_0) = (x_1, y_1)$;
- $\Phi(\cdot, x_0, y_0)$ solves (160) in the sense of Proposition 6.4.

Therefore every connected component of the intersection of the curve $\Phi(\cdot, x_0, y_0)$ with $[0, 1] \times [-\lambda, \lambda] \times [0, 1]$ is a straight segment lying on a plane $\{y = \text{const}\}$. If $(\tau, x_1, y_1) \in [0, 1]^3 \subset [0, 1] \times [-\lambda, \lambda] \times [0, 1]$, one of these segments contains (τ, x_1, y_1) and hence its slope is given by $f(\rho_n(\tau, x_1, y_1))$. If we choose λ large enough, the curve $\Phi(\cdot, x_0, y_0)$ remains trapped on the plane $\{y = y_1\}$ for the whole time interval $]0, \tau[$. Note that this choice of λ depends only on f and on the L^{∞} norm of ρ_n , which is uniformly bounded.

Fig. 2. The function $\theta_n(t, \cdot)$ and the segments $S_{j,i,l}$.

From now on, we assume that λ (and hence M) have been chosen so to satisfy the requirement above. Recall that for $\mathcal{L}^3$ -a.e. $(t, x, y) \in [0, 1]^3$, we have $\rho_n(t, x, y) = |c| + r_i$ for some i , and hence $f(\rho_n(t, x, y)) = 2^{-i}$. From the previous discussion we conclude the following formulas, valid for $\mathcal{L}^3$ -a.e. $(t, x, y) \in [0, 1]^3$:

- if $\bar{\rho}_n(x, y) = |c|$, then $\theta_n(t, x, y) = \bar{\theta}_n(x, y)$;
- if $\bar{\rho}_n(x, y) = |c| + r_i$, then $\theta_n(t, x, y) = \bar{\theta}_n(x - t2^{-i}, y)$.

Hence, for $j \in \{1, m_i - 1\}$, $i \geq n$, and $l \in \{1, \dots, 2^i - 1\}$, the function $\theta_n(t, \cdot)$ jumps on the segments

$$S_{j,i,l} := \left\{ y = 2^{-i} + \frac{j2^{-i}}{m_i} \mid x \in [l2^{-i}, (l+t)2^{-i}] \right\}.$$

See Figure 2.

The total amount of this jump is given by

$$J_i := \int_{S_{j,i,l}} |(\theta_n)^+(t, x, y) - (\theta_n)^-(t, x, y)| d\mathcal{H}^1(x) = t2^{-i} |\alpha_i - \alpha| = t2^{-i} i^{-2}.$$

Thus

$$\begin{aligned} \|\theta_n(t, \cdot)\|_{BV([0,1]^2)} &\geq \sum_{i \geq n} \sum_{j=1}^{m_i-1} \sum_{l=1}^{2^i-1} J_i = \sum_{i \geq n} (2^i - 1)(m_i - 1) J_i \\ &\geq \frac{t}{2} \sum_{i \geq n} (m_i - 1) i^{-2}. \end{aligned} \tag{161}$$

Clearly, since $|u_n|(t, \cdot) \in BV \cap L^\infty$ for every t and it is bounded away from zero, it is sufficient to show that $\theta_n(t, \cdot) \notin BV([0, 1]^2)$ for any $t \in]0, 1[$.

Recall that the bound (157) is the only condition required on the sequence of even numbers $\{m_i\}$. If we set $m_i = 2i^2$, then (157) is clearly satisfied, whereas (161) is infinite. $\square$

6.3. Proof of Proposition 6.5

As in the proof of Proposition 6.6, for $\beta \in \mathbb{R}$ we denote by $[\beta]$ the largest integer which is less than or equal to β .

The idea is to mimic the construction of Proposition 6.6. Hence we want to start with piecewise constant initial moduli $\bar{\rho}_n$ which are constant along $m - 1$ orthogonal directions $e_1, \dots, e_{m-1}$ and oscillate along the direction ω orthogonal to each e_i . The solution ρ_n of the scalar law (117) will then be constant along the directions $e_1, \dots, e_{m-1}$. Moreover, for small times, this solution will consist of shocks and rarefaction waves which do not interact. We will impose two requirements on this construction:

- We choose ω and the sizes and heights of the oscillations in such a way that the distinct shocks and rarefaction waves do not interact for times less than 1. Hence, in this range of times, between each couple of nearby shock and rarefaction wave, there will be a spacetime strip on which ρ is constant (see Figure 3).
- We choose ω in such a way that the trajectories of solutions of (150) are trapped in the strips for a sufficiently long time.

Finally we choose initial data $\bar{\theta}_n$ which oscillate along a direction perpendicular to ω , in such a way that in the strip mentioned above θ_n reproduce the behavior of the construction of Proposition 6.6.

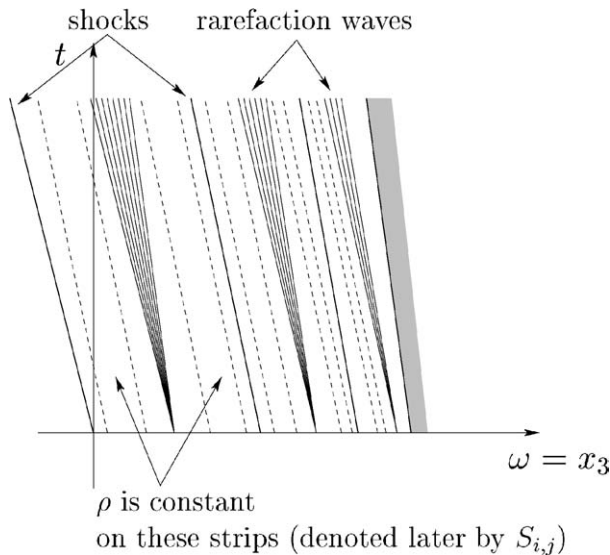


Fig. 3. A (t, ω) -slice of the evolution of ρ_n .

These requirements translate into geometric conditions on ω and into analytical ones on the various parameters which govern the oscillations. When $m \geq 3$ and g is not constant we can always satisfy these conditions. When $m = 2$, we are able to do it only in some cases.

Since the construction is the same, we only present the proof when $m \geq 3$ and, without loosing our generality, we assume $m = 3$. We denote by h the function given by $h(\rho) = \rho g(\rho)$ and by β the positive real number $|c|$. Clearly there exists a unit vector $\omega \in \mathbb{R}^3$ such that

$$\omega \cdot g(\beta) = \omega \cdot h'(\beta), \quad (162)$$

$$\omega \cdot g'(\beta) = 0, \quad (163)$$

$$\omega \cdot h''(\beta) = 0. \quad (164)$$

Indeed, since $h'(\beta) = g(\beta) + \beta g'(\beta)$, (162) reduces to (163). Thus, the conditions above reduce to find a unit vector $\omega \in \mathbb{R}^3$ which is perpendicular to both the vectors $g'(\beta)$ and $h''(\beta)$. We fix an orthonormal system of coordinates in $\mathbb{R}^3$ in such a way that $\omega = (0, 0, 1)$.

Step 1: Construction of the modulus. Let $\{\sigma_l\}$ be a sequence of vanishing positive real numbers such that $\sum \sigma_l < \infty$ and let $I_l \subset \mathbb{R}$ be the intervals

$$I_1 := [0, \sigma_1[\quad I_l := \left[\sum_{i \leq l-1} \sigma_i, \sum_{i \leq l} \sigma_i \right].$$

Let m_l be a strictly increasing sequence of even integers and divide every I_l in m_l equal subintervals I_l^j for $j \in \{1, \dots, m_l\}$. Finally, let $\{a_l\}$ be a vanishing sequence of real numbers and set

$$\rho^{\text{in}}(x_1, x_2, x_3) := \begin{cases} \beta + a_l & \text{if } x_3 \in I_l^j \text{ for some even } j, \\ \beta & \text{otherwise.} \end{cases}$$

Then, let ρ be the entropy solution of the Cauchy problem

$$\begin{cases} \partial_t \rho + \text{div}_x [h(\rho)] = 0, \\ \rho(0, \cdot) = \rho^{\text{in}}. \end{cases} \quad (165)$$

Clearly ρ is a function of t and x_3 only. Moreover, recalling that $(h^3)''(\beta) = 0$, we can apply Lemma 6.2 in order to get the following property.

(T) For every $C_1 > 0$, there exists a $C_2 > 0$ such that if

$$\frac{\sigma_l}{m_l} \geq C_2 a_l^2, \quad (166)$$

then every I_l^j contains a subinterval J_l^j such that

- the length of J_l^j is greater than $C_1 a_l^2$;

- for every $(t, \xi_1, \xi_2, \xi_3) \in [0, 1] \times \mathbb{R}^2 \times J_l^j$ we have

$$\rho(t, \xi_1, \xi_2, \xi_3 + t(h^3)'(\beta)) = \rho(0, \xi_1, \xi_2, \xi_3). \quad (167)$$

For each couple j, l we let $S_{l,j}$ be the strip

$$S_{l,j} := \{(t, x_1, x_2, x_3) \mid 0 \leq t \leq 1 \text{ and } (x_3 - th_3'(\beta)) \in J_l^j\}.$$

Step 2: The flux generated by ρ . Denote by $B_R \subset \mathbb{R}^3$ the ball of radius R centered at the origin. It is easy to check that there exists a constant C_3 such that:

$$\|\rho^{\text{in}}\|_{BV(B_R)} \leq C_3 R^3 + C_3 R^2 \left(\sum_l (m_l + 1) |a_l| \right). \quad (168)$$

Hence, to insure that $\rho^{\text{in}} \in BV_{\text{loc}}$ it is sufficient to assume

$$\sum_l (m_l + 1) |a_l| < \infty. \quad (169)$$

Assuming that this condition is fulfilled, from the classical result of Kruzhkov we get the existence of a constant M such that $\|\rho\|_{BV([0,1] \times B_R)} \leq M \|\rho^{\text{in}}\|_{BV(B_{R+Mt})}$. Thus we can consider the regular Lagrangian flow Φ for the ODE

$$\begin{cases} \frac{d}{dt} \Phi(s, \cdot) = g(\rho(s, \Phi(s, \cdot))), \\ \Phi(0, x) = x \end{cases}$$

(see Propositions 6.3 and 6.4). Fix any strip $S_{l,j}$ as defined in Step 1. Clearly, for a.e. x , every connected component of the intersection of the trajectory curve $\gamma_x := \{\Phi(t, x) \mid t \in \mathbb{R}\}$ with the strip $S_{l,j}$ is a straight segment. If j is even, then this segment is parallel to $(1, g(\beta))$, otherwise it is parallel to $(1, g(\beta + a_l))$. Thus, if j is even and $(t, x) \in S_{l,j}$, then the portion of trajectory

$$T_{t,x} := \{\Phi(s, \xi) \text{ for } \xi \text{ such that } \Phi(t, \xi) = x \text{ and for } s \in [0, t]\}$$

is a straight segment contained in $S_{l,j}$.

Let us now turn to the case where j is odd. Note that

$$g(\beta + a_l) = g(\beta) + g'(\beta)a_l + O(a_l^2). \quad (170)$$

Thanks to the properties of $\omega = (0, 0, 1)$, we have that the segments of the form

$$\{(t, \xi + t(g(\beta) + a_l g'(\beta))) \mid 0 \leq t \leq 1 \text{ and } (0, \xi) \in S_{l,j}\} \quad (171)$$

are subsets of $S_{l,j}$. Recall (T) of Step 1. From (170) and (171) it follows that, for C_1 in (T) sufficiently large, there exists a subinterval $K_{l,j}$ such that

- the length of $K_{l,j}$ is greater than a_l^2 ;
- if $t \in [0, 1]$ and $x_3 - tg'(\beta) \in K_{l,j}$, then the set

$$T_{t,x} = \{\Phi(s, \xi) \mid s \in [0, t] \text{ and } \Phi(t, \xi) = x\}$$

is a straight segment contained in $S_{l,j}$.

From now on we pick a C_1 (and hence C_2) in such a way to ensure the existence of the segments $K_{l,j}$.

Step 3: Construction of the angular part. We recall that $g'_3(\beta) = g'(\beta) \cdot \omega = 0$ and that $g'_3(\beta) \neq 0$. Since the construction of the previous step is independent of the choice of the coordinates x_1 and x_2 , we can choose them so that $g'(\beta) = (0, C_4, 0)$, with $C_4 > 0$. Choose the a_l 's in such a way that

$$g_2(\beta + a_l) - g_2(\beta) = 2^{-l}.$$

Then, clearly, there exists a constant C_5 such that

$$\frac{2^{-l}}{C_5} \leq a_l \leq C_5 2^{-l}. \quad (172)$$

Set $\eta = c/|c|$ and let $\eta_l \in \mathbf{S}^{k-1}$ be such that $|\eta_l - \eta| = l^{-2}$. Then define

$$\theta^{\text{in}}(x_1, x_2, x_3) := \begin{cases} \eta_l & \text{if } x_3 \in I_l \text{ and } [2^l x_2] \text{ is even,} \\ \eta & \text{otherwise.} \end{cases}$$

Set $u^{\text{in}} := \rho^{\text{in}} \theta^{\text{in}}$. Let u be the renormalized entropy solution of

$$\begin{cases} \partial_t u + \operatorname{div}_z [g(|u|)u] = 0, \\ u(0, \cdot) = u^{\text{in}}. \end{cases} \quad (173)$$

We denote by θ the angular part $u/|u|$. According to Propositions 6.3 and 6.4, θ is given by the formula

$$\theta(t, x) = \theta^{\text{in}}(\Psi(t, x)),$$

where Ψ is a map such that $\Phi(t, \Psi(t, x)) = \Psi(t, \Phi(t, x)) = x$ for $\mathcal{L}^4$ -a.e. (t, x) . In what follows we denote by Φ_t^{-1} the map $\Psi(t, \cdot)$.

Step 4: Choice of parameters. We will prove that, for an appropriate choice of the various parameters, $u^{\text{in}} \in BV_{\text{loc}}$, whereas $u(t, \cdot)$ is not in BV_{loc} for any $t \in]0, 1]$. Recall that $\rho^{\text{in}} = |u^{\text{in}}|$ and $\rho(t, \cdot) = |u|(t, \cdot)$ are both in BV_{loc} and that $C_6^{-1} \leq \rho \leq C_6$ for some positive constant C_6 . Thus our goal is to choose the parameters σ_l and m_l in such a way that $\theta^{\text{in}} \in BV_{\text{loc}}$ and $\theta(t, \cdot) \notin BV_{\text{loc}}$ for every $t \in]0, 1]$. Note that, for some constant C_7 ,

$$\|\theta^{\text{in}}\|_{BV(B_R)} \leq C_7 R^3 + C_7 R^2 \left(\sum_l \frac{2^l}{l^2} \sigma_l + \sum_l l^{-2} \right). \quad (174)$$

Hence, choosing $\sigma_l = 2^{-l}$ we conclude that $\theta^{\text{in}} \in BV(B_R)$ for every $R > 0$.

Now, we choose $m_l = 2l^2$, and since from (172) we have $a_l \leq C_5^2 2^{-l}$, we clearly fulfil the condition (169), which is the only one we required on the sequence $\{m_l\}$. Thus we get

$$\frac{\sigma_l}{m_l} = l^{-2} 2^{-l+1}.$$

Since from (172) we have $a_l^2 \leq C_5 2^{-2l}$, clearly (166) is fulfilled for any constant C_2 , provided l is large enough. Thus, we get the existence of a constant C_8 such that the segments $K_{l,j}$ of Step 2 exist for any $l \geq C_8$.

Fix $t \in]0, 1]$ and $l \geq C_8$. Recalling that $\theta(t, x) = \theta^{\text{in}}(\Phi_t^{-1}(x))$ and taking into account the properties of Φ proved in the Step 2, we conclude what follows.

- If $j \in [1, m_l]$ is even and $\xi_{l,j}$ belongs to the segment $J_{l,j}$, then

$$\theta(t, x_1, x_2, \xi_{l,j} + tg^3(\beta)) = \begin{cases} \eta_l & \text{if } [2^l(x_2 - tg^2(\beta))] \text{ is even,} \\ \eta & \text{otherwise.} \end{cases}$$

- If $j \in [1, m_l]$ is odd and $\xi_{l,j}$ belongs to the segment $K_{l,j}$, then

$$\theta(t, x_1, x_2, \xi_{l,j} + tg^3(\beta)) = \begin{cases} \eta_l & \text{if } [2^l(x_2 - tg^2(\beta + a_l))] \text{ is even,} \\ \eta & \text{otherwise.} \end{cases}$$

Recall that $g_2(\beta + a_l) - g_2(\beta) = 2^{-l}$. Thus, for any $j \in [1, m_l - 1]$, we have

$$\begin{aligned} A_{l,j} &:= \int_{[0,1]^2} |\theta(t, x_1, x_2, \xi_{l,j} + tg^3(\beta)) - \theta(t, x_1, x_2, \xi_{l,j+1} + tg^3(\beta))| dx_1 dx_2 \\ &= t |\eta_l - \eta| = tl^{-2}. \end{aligned}$$

Thus

$$\sum_{l \geq C_8} \sum_{1 \leq j \leq m_l - 1} A_{l,j} = t \sum_{l \geq C_8} \frac{m_l - 1}{l^2} = t \sum_{l \geq C_8} \frac{2l^2 - 1}{l^2} = \infty. \quad (175)$$

Note that if $\theta(t, \cdot)$ were locally in BV , then $\partial_{x_3} \theta(t, \cdot)$ would be a Radon measure. Denote by μ the total variation measure of $\partial_{x_3} \theta(t, \cdot)$ and by $\mathcal{S}_{l,j}$ the stripes

$$\mathcal{S}_{l,j} := \{(x_1, x_2, x_3) \mid (x_1, x_2) \in [0, 1]^2 \text{ and } (x_3 - tg^3(\beta)) \in [\xi_{l,j}, \xi_{l,j+1}]\}.$$

Then $A_{l,j} \leq \mu(\mathcal{S}_{l,j})$. The $\mathcal{S}_{l,j}$ are pairwise disjoint and for R' sufficiently large, they are all contained in the ball $B_{R'}$. Thus, we would get

$$\sum_{l \geq C_8} \sum_{1 \leq j \leq m_l - 1} A_{l,j} \leq \sum_{l \geq C_8} \sum_{1 \leq j \leq m_l - 1} \mu(\mathcal{S}_{l,j}) \leq \mu(B_{R'}) < \infty,$$

which contradicts (175). Hence, we conclude that $\theta(t, \cdot)$ is not in $BV(B_{R'})$ for any $t \in]0, 1]$.

Step 5. Truncation of the construction and conclusion. Next, define $\hat{u}_n^{\text{in}}: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ as

$$\hat{u}_n^{\text{in}}(x_1, x_2, x_3) := \begin{cases} u^{\text{in}}(x_1, x_2, x_3) & \text{if } x_3 \in I_l \text{ for some } l \geq n, \\ c & \text{otherwise.} \end{cases}$$

Clearly $\|\hat{u}_n^{\text{in}} - c\|_\infty + \|\hat{u}_n^{\text{in}} - c\|_{BV(\Omega)} \rightarrow 0$ for every bounded open set $\Omega \subset \mathbb{R}^3$. Moreover, if we denote by $\hat{u}_n$ the renormalized entropy solution of

$$\begin{cases} \partial_t u + \operatorname{div}_z [g(|u|)u] = 0, \\ u(0, \cdot) = \hat{u}_n^{\text{in}}, \end{cases} \quad (176)$$

then $\hat{u}_n(t, \cdot) \notin BV(B_{R'})$ for any $t \in]0, 1]$. Finally, let $M > 0$ and define

$$\bar{u}_n(x_1, x_2, x_3) := \begin{cases} \hat{u}_n^{\text{in}}(x_1, x_2, x_3) & \text{if } x_1^2 + x_2^2 + x_3^2 \leq M, \\ c & \text{otherwise.} \end{cases}$$

Let u_n be the renormalized entropy solution of

$$\begin{cases} \partial_t u + \operatorname{div}_x [g(|u|)u] = 0, \\ u(0, \cdot) = \bar{u}_n. \end{cases} \quad (177)$$

For any $M' > 0$, by the finite speed of propagation for scalar laws, if we choose M sufficiently large, then $|u_n| = |\hat{u}_n|$ on $[0, 1] \times B_{M'}(0)$. Using Lemma 3.17 and arguing as in the proof of Theorem 6.1, we conclude that $u_n = \hat{u}_n$ on $[0, 1] \times B_{R'}(0)$, provided M' is chosen sufficiently large.

7. Partial regularity and trace properties of solutions to transport equations

In this chapter we will show two regularity properties of solutions to transport equations proved in [6]. The first one is a trace property. Namely, if

- B is a bounded BV vector field and μ a Radon measure,
- w is a bounded solution of the equation

$$D \cdot (wB) = \mu, \quad (178)$$

• and Σ is a noncharacteristic hypersurface for (178),
then w has a strong L^1 trace on Σ .

More precisely:

THEOREM 7.1. *Let B be a bounded BV vector field in $\Omega \subset \mathbb{R}^d$ and w an L^∞ function such that $D \cdot (wB)$ is a Radon measure. Let Σ be an oriented C^1 hypersurface with normal ν such that $\nu \cdot B^+ \neq 0$ $\mathcal{H}^{d-1}$ -a.e. on Σ . Then for $\mathcal{H}^{d-1}$ -a.e. $x \in \Sigma$ there exists $w^+(x) \in \mathbb{R}$ such that*

$$\lim_{r \downarrow 0} \frac{1}{r^d} \int_{B^+(x, \nu)} |w(y) - w^+(x)| \, dy = 0. \quad (179)$$

REMARK 7.2. In [6] the authors proved this result for the larger class of vector fields B of bounded deformation. The proof of this stronger result is not substantially different but it needs some adjustments, which go beyond the aims of these notes.

The second property concerns Lebesgue points of w . Before stating it let us introduce the tangential set of a BV vector field.

DEFINITION 7.3 (Tangential set of B). Let $B \in BV_{\text{loc}}(\Omega, \mathbb{R}^d)$, let $|DB|$ denote the total variation of its distributional derivative and denote by $\tilde{E}$ the Borel set of points $x \in \Omega$ such that

- the following limit exists and is finite

$$M(x) := \lim_{r \downarrow 0} \frac{DB(B_r(x))}{|DB|(B_r(x))};$$

- the Lebesgue limit $\tilde{B}(x)$ exists.

We call tangential set of B the Borel set

$$E := \{x \in \tilde{E} \text{ such that } M(x) \cdot \tilde{B}(x) = 0\}.$$

THEOREM 7.4. Let $B \in BV_{\text{loc}}(\Omega, \mathbb{R}^d)$ and let $w \in L_{\text{loc}}^\infty(\Omega)$ be such that $D \cdot (Bw)$ is a locally finite Radon measure in Ω . Then $|D^c B|$ -a.e. point $x \notin E$ is a Lebesgue point for w , and hence for any such x there exists $\tilde{w}(x)$ such that

$$\lim_{r \downarrow 0} \frac{1}{r^n} \int_{B_r(x)} |w(y) - \tilde{w}(x)| \, dy = 0. \quad (180)$$

The proof of this theorem relies on Theorem 7.1, on the Alberti's rank-one theorem (Theorem 2.13) and on the coarea formula.

7.1. Anzellotti's weak trace for measure – divergence bounded vector fields

In this section we recall some basic facts about the trace properties of vector fields whose divergence is a measure (see [12], the unpublished work [13,14,20,21], and finally [6]).

Thus, let $U \in L_{\text{loc}}^\infty(\Omega, \mathbb{R}^d)$ be such that its distributional divergence $D \cdot U$ is a measure with locally finite variation in Ω . The starting point is to define for every C^1 open set $\Omega' \subset \Omega$ the distribution $\text{Tr}(U, \partial\Omega')$ as

$$\langle \text{Tr}(U, \partial\Omega'), \varphi \rangle := \int_{\Omega'} \nabla \varphi \cdot U + \int_{\Omega'} \varphi \, d[D \cdot U] \quad \forall \varphi \in C_c^\infty(\Omega). \quad (181)$$

It was proved in [12] the following proposition.

PROPOSITION 7.5. *There exists a unique $g \in L^\infty_{\text{loc}}(\Omega \cap \partial\Omega')$ such that*

$$\langle \text{Tr}(U, \partial\Omega'), \varphi \rangle = \int_{\partial\Omega'} g \varphi \, d\mathcal{H}^{d-1}.$$

PROOF. Clearly, the support of the distribution $\text{Tr}(U, \partial\Omega')$ is contained in $\partial\Omega'$.

Next we claim that for any $\varphi \in C_c^\infty(\Omega)$ and any $\varepsilon > 0$ there exists $\hat{\varphi}_\varepsilon \in C_c^\infty(\Omega)$ such that

- (i) $\hat{\varphi}_\varepsilon - \varphi_\varepsilon$ vanishes in a neighborhood of $\partial\Omega'$;
- (ii) $\|\hat{\varphi}_\varepsilon\|_\infty \leq \|\varphi\|_\infty$;
- (iii) $\hat{\varphi}_\varepsilon = 0$ on $\Omega'_\varepsilon := \{x \in \Omega' : \text{dist}(x, \partial\Omega') > \varepsilon\}$;
- (iv) $\int_{\Omega'} |\nabla \hat{\varphi}_\varepsilon| \leq \varepsilon + \int_{\partial\Omega'} |\varphi|$.

Having such a $\hat{\varphi}_\varepsilon$ we can easily estimate

$$\begin{aligned} |\langle \text{Tr}(U, \partial\Omega'), \varphi \rangle| &= |\langle \text{Tr}(U, \partial\Omega'), \hat{\varphi}_\varepsilon \rangle| \\ &\leq \left| \int_{\Omega'} \hat{\varphi}_\varepsilon \, d[D \cdot U] \right| + \|U\|_{L^\infty(\Omega')} \int_{\Omega'} |\nabla \hat{\varphi}_\varepsilon| \\ &\leq \int_{\Omega' \setminus \Omega'_\varepsilon} |\hat{\varphi}_\varepsilon| \, d[D \cdot U] + \|U\|_{L^\infty(\Omega')} \left(\int_{\partial\Omega'} |\varphi| + \varepsilon \right) \\ &\leq \|\varphi\|_\infty |D \cdot U|(\Omega' \setminus \Omega'_\varepsilon) + \|U\|_{L^\infty(\Omega')} \left(\int_{\partial\Omega'} |\varphi| + \varepsilon \right). \end{aligned}$$

Letting $\varepsilon \downarrow 0$ we get $|\langle \text{Tr}(U, \partial\Omega'), \varphi \rangle| \leq \|U\|_\infty \|\varphi\|_{L^1(\partial\Omega')}$. This estimate is valid for any $\varphi \in C_c^\infty(\Omega)$ and therefore implies the claim of the proposition.

It remains to prove the existence of the function $\hat{\varphi}_\varepsilon$. Using the fact that $\partial\Omega'$ is locally the graph of a C^1 function, we can find a family of open sets $\{\Omega_h\}_{h \in \mathbb{N}}$ such that $\Omega_h \subset \subset \Omega$, $\Omega_h \uparrow \Omega'$ and

$$\limsup_{h \uparrow \infty} |D\mathbb{1}_{\Omega_h}|(\mathbb{R}^d) \leq |D\mathbb{1}_{\Omega'}|(\mathbb{R}^d).$$

Let $\varphi \in C_c^\infty(\Omega)$ and $\varepsilon > 0$ be given and consider h so large that $\overline{\Omega'_\varepsilon} \subset \Omega_h$. Let $\{\eta_\delta\}_{\delta > 0}$ be a standard family of mollifiers and choose $\delta = \delta(h) < \text{dist}(\partial\Omega', \partial\Omega_h)$ so small that $\Omega'_\varepsilon \subset \{\mathbb{1}_{\Omega_h} * \eta_\delta(h) = 1\}$. Set $\zeta_h := \mathbb{1}_{\Omega_h} * \eta_\delta(h)$ and $\hat{\varphi}_\varepsilon := \varphi(1 - \zeta_h)$. Clearly $\hat{\varphi}_\varepsilon$ satisfies (i)–(iii). Therefore it remains to check that (iv) holds for h sufficiently large. Indeed, note that

$$\limsup_{h \uparrow \infty} \int |\nabla \zeta_h| \leq |D\mathbb{1}_{\Omega'}|(\mathbb{R}^d).$$

Since $\zeta_h \rightarrow \mathbb{1}_{\Omega'}$ in L^1 , for every open set A , we get

$$\liminf_{h \uparrow \infty} \int_A |\nabla \zeta_h| \geq |D\mathbb{1}_{\Omega'}|(A)$$

for every open set A . Therefore we conclude that the measures $|\nabla \zeta| \mathcal{L}^d$ converges weakly* to $\mathcal{H}^{d-1} \llcorner \partial \Omega'$. Hence we have

$$\int_{\Omega'} |\nabla \hat{\varphi}_\varepsilon| \leq \int_{\Omega'} (1 - \zeta_h) |\nabla \varphi| + \int_{\Omega'} |\varphi| |\nabla \zeta_h| \rightarrow \int_{\partial \Omega'} |\varphi| d\mathcal{H}^{d-1}.$$

This shows that (iv) holds for h sufficiently large, and thus completes the proof of the proposition. $\square$

By a slight abuse of notation, we denote the function g by $\text{Tr}(U, \partial \Omega')$ as well.

REMARK 7.6. Clearly the notion of trace is local, that is, if $A \subset \partial \Omega_1 \cap \partial \Omega_2$ is relatively open and the outer normals of $\partial \Omega_1$ and $\partial \Omega_2$ coincide on Σ , then $\text{Tr}(U, \partial \Omega_1) = \text{Tr}(U, \partial \Omega_2)$ on Σ .

Given an oriented C^1 hypersurface Σ , we can always view it locally as the boundary of an open set Ω_1 having ν_Σ as unit exterior normal. In this way, we can define the positive trace $\text{Tr}^+(U, \Sigma)$ as $\text{Tr}(U, \partial \Omega_1)$ and the negative trace $\text{Tr}^-(U, \Sigma)$ as $-\text{Tr}(U, \Omega_2 \setminus \Omega_1)$, where Ω_2 is any open set such that $\Omega_1 \subset \subset \Omega_2 \subset \subset \Omega$. The locality property of Remark 7.6 gives that both $\text{Tr}^-(U, \Sigma)$ and $\text{Tr}^+(U, \Sigma)$ are well defined.

In order to extend the notion of trace to countably $\mathcal{H}^{d-1}$ -rectifiable sets, we need a stronger locality property: In [12] it was proved the following proposition.

PROPOSITION 7.7. *If $\Omega_1, \Omega_2 \subset \subset \Omega$ are two C^1 open sets, then*

$$\text{Tr}(U, \partial \Omega_1) = \text{Tr}(U, \partial \Omega_2) \quad \mathcal{H}^{d-1}\text{-a.e. on } \partial \Omega_1 \cap \partial \Omega_2, \quad (182)$$

if the exterior unit normals coincide on $\partial \Omega_1 \cap \partial \Omega_2$.

Here we follow the recent proof of [6].

PROOF. Set $\mu := |D \cdot U| \llcorner \Omega_1 \cup \Omega_2$ and $E := \partial \Omega_1 \cap \partial \Omega_2$, and denote by T_i the $L^\infty(\partial \Omega_i)$ function which gives the trace $\text{Tr}(U, \partial \Omega_i)$. Note that from our assumptions it follows that $\mu(E) = 0$. This implies that

- (i) $\mu(B_r(x)) = o(r^{d-1})$ for $\mathcal{H}^{d-1}$ -a.e. $x \in E$ (see, for instance, Theorem 2.53 of [11]);
- (ii) $\mathcal{H}^{d-1}$ -a.e. $x \in E$ is a Lebesgue point for T_1 and T_2 .

It suffices to show $T_1(x) = T_2(x)$ for any x satisfying both (i) and (ii).

Thus, let x be any such point and χ_r a test function $\chi \in C_c^\infty(B_1(0))$ with $0 \leq \chi \leq 1$. Set $\chi_r(y) := \chi((y-x)/r)$ for every positive r . When r is small enough, we get $\text{supp}(\chi_r) \subset \Omega$ and thus

$$\int_{\partial \Omega_i} T_i \chi_r = \int_{\Omega_i} \nabla \chi_r \cdot U + \int_{\Omega_i} \chi_r d[D \cdot U].$$

Hence

$$\begin{aligned} & \left| \int_{\partial\Omega_1} T_1 \chi_r - \int_{\partial\Omega_2} T_2 \chi_r \right| \\ & \leq \left| \int_{\Omega_1} \nabla \chi_r \cdot U - \int_{\Omega_2} \nabla \chi_r \cdot U \right| + \left| \int_{\Omega_1} \chi_r d[D \cdot U] - \int_{\Omega_2} \chi_r d[D \cdot U] \right|. \end{aligned}$$

Note that, since x is a Lebesgue point for both $T_i \tilde{\Omega}$ s, for some constant C_χ (depending only on χ) we have

$$\lim_{\rho \downarrow 0} \frac{1}{r^{d-1}} \left| \int_{\partial\Omega_1} T_1 \chi_r - \int_{\partial\Omega_2} T_2 \chi_r \right| = C_\chi |T_1(x) - T_2(x)|. \quad (183)$$

Moreover, C_χ is positive if, for instance, $\chi = 1$ on $B_{1/2}(0)$. Therefore it suffices to show that

$$\lim_{\rho \downarrow 0} \frac{1}{r^{d-1}} \left| \int_{\Omega_1} \nabla \chi_r \cdot U - \int_{\Omega_2} \nabla \chi_r \cdot U \right| = 0 \quad (184)$$

and

$$\lim_{\rho \downarrow 0} \frac{1}{r^{d-1}} \left| \int_{\Omega_1} \chi_r d[D \cdot U] - \int_{\Omega_2} \chi_r d[D \cdot U] \right| = 0 \quad (185)$$

to conclude that the right-hand side of (183) vanishes and $T_1(x) = T_2(x)$.

Since $|\nabla \chi_r| \geq C/r$, we have

$$\begin{aligned} \left| \int_{\Omega_1} U \cdot \nabla \chi_r - \int_{\Omega_2} U \cdot \nabla \chi_r \right| & \leq \frac{C}{r} \mathcal{L}^d((\Omega_1 \setminus \Omega_2 \cup \Omega_2 \setminus \Omega_1) \cap B_r(x)) \\ & = o(r^{d-1}), \end{aligned}$$

which shows (184).

On the other hand

$$\begin{aligned} & \left| \int_{\Omega_1} \chi_r d[D \cdot U] - \int_{\Omega_2} \chi_r d[D \cdot U] \right| \\ & \leq \|\chi_r\|_\infty |D \cdot U|((\Omega_1 \setminus \Omega_2 \cup \Omega_2 \setminus \Omega_1) \cap B_r(x)) \\ & \leq \mu(B_r(x)) = o(r^{d-1}), \end{aligned}$$

which implies (185). □

Using the decomposition of a rectifiable set Σ in pieces of C^1 hypersurfaces we can define an orientation of Σ and the normal traces of U on Σ as follows:

DEFINITION 7.8. By the rectifiability property we can find countably many oriented C^1 hypersurfaces Σ_i and pairwise disjoint Borel sets $E_i \subset \Sigma_i \cap \Sigma$ such that $\mathcal{H}^{d-1}(\Sigma \setminus \bigcup_i E_i) = 0$; then we define $\nu_\Sigma(x)$ equal to the classical normal to Σ_i for any $x \in E_i$. Analogously, we define

$$\mathrm{Tr}^+(U, \Sigma) := \mathrm{Tr}^+(U, \Sigma_i), \quad \mathrm{Tr}^-(U, \Sigma) := \mathrm{Tr}^-(U, \Sigma_i) \quad \mathcal{H}^{d-1}\text{-a.e. on } E_i.$$

The locality property of Proposition 7.7 ensures that this definition depends on the orientation ν_Σ , as in the case of oriented C^1 hypersurfaces, but, up to $\mathcal{H}^{d-1}$ -negligible sets, it does not depend on the choice of Σ_i and E_i .

7.2. Further properties of Anzellotti's weak trace

In this section we follow [6] and collect three important properties of the trace of bounded vector fields with measure divergence.

PROPOSITION 7.9 (Jump part of $D \cdot U$). *Let the divergence of $U \in L^\infty_{\mathrm{loc}}(\Omega, \mathbb{R}^d)$ be a measure with locally finite variation in Ω . Then*

- (a) $|D \cdot U|(E) = 0$ for any $\mathcal{H}^{d-1}$ -negligible set $E \subset \Omega$.
- (b) If $\Sigma \subset \Omega$ is a C^1 hypersurface then

$$D \cdot U \llcorner \Sigma = (\mathrm{Tr}^+(U, \Sigma) - \mathrm{Tr}^-(U, \Sigma)) \mathcal{H}^{d-1} \llcorner \Sigma. \quad (186)$$

Thanks to Proposition 7.9(a) it turns out that for any $U \in L^\infty_{\mathrm{loc}}(\Omega, \mathbb{R}^d)$ whose divergence is a locally finite measure in Ω there exist a Borel function f and a set $J = J_{D \cdot U}$ such that

$$D^j \cdot U = f \mathcal{H}^{d-1} \llcorner J_{D \cdot U}. \quad (187)$$

PROPOSITION 7.10 (Fubini's theorem for traces). *Let U be as above and let $F \in C^1(\Omega)$. Then*

$$\mathrm{Tr}(U, \partial\{F > t\}) = U \cdot \nu \quad \mathcal{H}^{d-1}\text{-a.e. on } \Omega \cap \partial\{F > t\}$$

for $\mathcal{L}^1$ -a.e. $t \in \mathbb{R}$, where ν denotes the exterior unit normal to $\{F > t\}$.

Notice that the coarea formula gives $\mathcal{H}^{d-1}(\{F = t\} \cap \{|\nabla F| = 0\}) = 0$ for $\mathcal{L}^1$ -a.e. $t \in \mathbb{R}$. Therefore the theory of traces applies to the sets $\Sigma_t = \{F = t\}$ for $\mathcal{L}^1$ -a.e. $t \in \mathbb{R}$.

THEOREM 7.11 (Weak continuity of traces). *Let $U \in L^\infty(\Omega, \mathbb{R}^d)$ be such that $D \cdot U$ is a Radon measure and let $f \in C^1(\mathbb{R}^{d-1})$. For $t \in \mathbb{R}$ consider the surfaces*

$$\Sigma_t := \{x: x_d = t + f(x_1, \dots, x_{d-1})\} \cap \Omega$$

and set

$$\alpha_t(x_1, \dots, x_{d-1}) := \text{Tr}(U, \Sigma_t)(x_1, \dots, x_{d-1}, f(x_1, \dots, x_{d-1}) + t).$$

If $D \subset \mathbb{R}^{d-1}$ is an open set and $I \subset \mathbb{R}$ an interval such that $\Omega' := \{(x', f(x') + t) : (x', t) \in D \times I\} \subset \Omega$, then for every $t_0 \in I$ we have $\alpha_t \rightharpoonup^* \alpha_{t_0}$ in $L^\infty(D)$ as $t \rightarrow t_0$.

PROOF OF PROPOSITION 7.9. Claim (a) has been proved in Lemma 2.4. Concerning claim (b), by the locality of the statement it suffices to prove that, if $A \subset\subset \Omega$ and $F \in C^1(A)$ are such that $\Sigma \cap A = \{F = 0\}$ and $\nabla F \neq 0$ on A , then

$$\begin{aligned} \int_{\Sigma} \varphi \, d[D \cdot U] \\ = \int_{\Sigma} \varphi [\text{Tr}(U, \partial\{F > 0\}) + \text{Tr}(U, \partial\{F < 0\})] \quad \text{for every } \varphi \in C_c^\infty(A). \end{aligned}$$

Note that

$$\begin{aligned} \int_{\Sigma} \varphi \, d[D \cdot U] &= \int_A \varphi \, d[D \cdot U] - \int_{\{F > 0\}} \varphi \, d[D \cdot U] - \int_{\{F < 0\}} \varphi \, d[D \cdot U] \\ &= - \int_A \nabla \varphi \cdot U + \int_{\{F > 0\}} \nabla \varphi \cdot U + \int_{\Sigma} \varphi \, \text{Tr}(U, \partial\{F > 0\}) \\ &\quad + \int_{\{F < 0\}} \nabla \varphi \cdot U + \int_{\Sigma} \varphi \, \text{Tr}(U, \partial\{F < 0\}) \\ &= \int_{\Sigma} \varphi [\text{Tr}(U, \partial\{F > 0\}) + \text{Tr}(U, \partial\{F < 0\})] \varphi. \quad \square \end{aligned}$$

PROOF OF PROPOSITION 7.10. The statement of the proposition is trivial if U is smooth. In the general case we will prove it by approximation.

Indeed let U be a field as in the statement of the proposition, choose a standard family of mollifiers $\{\eta_\varepsilon\}_{\varepsilon > 0}$ and set $U_\varepsilon := U * \eta_\varepsilon$. Recall that $|D \cdot U_\varepsilon| \rightharpoonup^* |D \cdot U|$ in the sense of measures. Note that the set $S := \{t : |D \cdot U|(\Sigma_t) = 0\}$ is at most countable. For $t \notin S$ we have

- $\text{Tr}^+(U, \Sigma_t) = \text{Tr}^-(U, \Sigma_t)$ by Proposition 7.9;
- $(D \cdot U_\varepsilon) \llcorner \{F > t\} \rightharpoonup^* (D \cdot U) \llcorner \{F > t\}$ and $(D \cdot U_\varepsilon) \llcorner \{F < t\} \rightharpoonup^* (D \cdot U) \llcorner \{F < t\}$ by Proposition 2.1.

Therefore, from the definition of trace it follows that

$$\text{Tr}(U_\varepsilon, \partial\{F > t\}) \rightharpoonup \text{Tr}(U, \partial\{F > t\})$$

in the sense of distributions for every $t \notin S$.

Since U_ε is smooth, $\text{Tr}(U_\varepsilon, \partial\{F > t\}) = U_\varepsilon \cdot \nu_t$ and therefore it suffices to prove that

- There exists a vanishing sequence $\{\varepsilon_h\}_{h \in \mathbb{N}} \subset \mathbb{R}^+$ such that

$$U_{\varepsilon_h} \cdot \nu \rightarrow U \cdot \nu \quad \text{in } L^1(\Sigma_t)$$

for $\mathcal{L}^1$ -a.e. t .

Such a property holds for every fast converging subsequence $\{U_{\varepsilon_h}\}$, i.e., such that

$$\sum_{h=1}^{\infty} \|U_{\varepsilon_h} - U\|_{L^1(\Omega)} < \infty.$$

Indeed for such a subsequence we can use the coarea formula to estimate

$$\begin{aligned} \int_{\mathbb{R}} \sum_h \|U_{\varepsilon_h} - U\|_{L^1(\Sigma_t)} dt &\leq \sum_h \int_{\mathbb{R}} \|U_{\varepsilon_h} - U\|_{L^1(\Sigma_t)} dt \\ &\leq \sum_h \int_{\Omega} |\nabla F| |U_{\varepsilon_h} - U| \\ &\leq \|F\|_{C^1} \sum_h \|U_{\varepsilon_h} - U\|_{L^1(\Omega)} < \infty. \end{aligned}$$

Thus, for $\mathcal{L}^1$ -a.e. t the series $\sum_h \|U_{\varepsilon_h} - U\|_{L^1(\Sigma_t)}$ must be finite, and this implies that for any such t , $U_{\varepsilon_h} \rightarrow U$ strongly in $L^1(\Sigma_t)$. $\square$

PROOF OF THEOREM 7.11. Let $\varphi \in C_c^\infty(D)$ be given and consider the function $\psi \in C^1(\Omega')$ given by $\psi(x', x_d) = \varphi(x')$. It is not difficult to see that ψ can be extended to a function in $C_c^1(\Omega)$. Next, set $\sigma(x') := \sqrt{1 + |\nabla f(x')|^2}$ and for every $t > t_0$ define the open set

$$\Omega_t := \{(x', f(x') + \tau) : x' \in D, \tau \in]t_0, t[\}.$$

In analogous way we define Ω_t for $t < t_0$. Then, using the definition of trace, we easily get

$$\begin{aligned} &\left| \int_D \varphi(x') \sigma(x') (\alpha_t(x') - \alpha_{t_0}(x')) dx' \right| \\ &= \left| \int_{\partial \Omega_t} (\text{Tr}^+(U, \Sigma_t)(x) \psi(x) - \text{Tr}^-(U, \Sigma_{t_0})(x) \psi(x)) d\mathcal{H}^{d-1}(x) \right| \\ &= \left| - \int_{\Omega_t} \nabla \psi \cdot U - \int_{\Omega_t} \psi d[D \cdot U] \right| \\ &\leq \|\nabla \psi\|_{L^\infty(\Omega_t)} \|U\|_\infty |\Omega_t| + \|\Phi\|_\infty |D \cdot U|(\Omega_t). \end{aligned}$$

Since the last expressions converge to 0 as $t \rightarrow t_0$, we get that

$$\int_D \varphi(x') \sigma(x') (\alpha_t(x') - \alpha_{t_0}(x')) dx' \rightarrow 0$$

for every $\varphi \in C_c^\infty(D)$. Since $\|\alpha_t\|_\infty$ is bounded by $\|U\|_\infty$, we conclude that $\alpha_t \sigma$ converges weakly* in $L^\infty(D)$ to $\alpha_{t_0} \sigma$. Note that $\sigma \geq 1$, and hence $\alpha_t \rightharpoonup^* \alpha_{t_0}$, which is the desired conclusion. $\square$

7.3. Change of variables for traces

This section is devoted to prove the core result of [6], namely the following Òchain ruleÓ for traces.

THEOREM 7.12 (Change of variables for traces). *Let $B \in BV \cap L^\infty(\Omega, \mathbb{R}^d)$ and $w \in L^\infty(\Omega)$ be such that $D \cdot (wB)$ is a Radon measure. If $\Omega' \subset \subset \Omega$ is an open domain with a C^1 boundary and $h \in C^1(\mathbb{R}^k)$, then*

$$\text{Tr}(h(w)B, \partial\Omega') = h\left(\frac{\text{Tr}(wB, \partial\Omega')}{\text{Tr}(B, \partial\Omega')}\right) \text{Tr}(B, \partial\Omega') \quad \mathcal{H}^{d-1}\text{-a.e. on } \partial\Omega'.$$

Here we use the convention that when $\text{Tr}(B, \partial\Omega')(x) = 0$, the expression

$$h\left(\frac{\text{Tr}(wB, \partial\Omega')(x)}{\text{Tr}(B, \partial\Omega')(x)}\right) \text{Tr}(B, \partial\Omega')(x)$$

is zero as well.

REMARK 7.13. In [6] the authors proved the previous theorem for the class of vector fields B of bounded deformation (compare with Remark 7.2).

In order to prove the theorem, we need the following renormalization lemma.

LEMMA 7.14. *Let B , w and h be as above. Then $D \cdot (h(w)B)$ is a Radon measure and, if $R := \|w\|_\infty$, then*

$$\begin{aligned} |D \cdot (h(w)B)| &\leq \|\nabla h\|_{L^\infty(B_R(0))} (|D \cdot (wB)| + 2R|D^s \cdot B|) \\ &\quad + \left(\sup_{v \in B_R(0)} \left| h(v) - \sum v^i \frac{\partial h}{\partial v^i}(v) \right| \right) |D \cdot B|. \end{aligned}$$

PROOF. Let $\{\eta_\delta\}$ be a family of standard mollifiers and set $w_\delta := w * \eta_\delta$, $T_\delta := (D \cdot (wB)) * \eta_\delta - D \cdot (w_\delta B)$. Then we compute

$$\begin{aligned} D \cdot h(w_\delta) &= \sum_i \frac{\partial h}{\partial v^i}(w_\delta) (D \cdot (B w^i)) * \eta_\delta + \sum_i \frac{\partial h}{\partial v^i}(w_\delta) T_\delta^i \\ &\quad + \left(h(w_\delta) - \sum_i \frac{\partial h}{\partial v^i}(w_\delta) w_\delta^i \right) D \cdot B. \end{aligned}$$

Using the commutator estimate of Proposition 4.6 and Lemma 4.8 we easily conclude (compare with the proof of Theorem 4.1). $\square$

PROOF OF THEOREM 7.12. It is not restrictive to assume that the larger open set Ω is bounded and it has a C^1 boundary.

Step 1. Let $\Omega'' = \Omega \setminus \overline{\Omega}'$. In this step we prove that

$$\mathrm{Tr}(h(w)B, \partial\Omega'') = h\left(\frac{\mathrm{Tr}(wB, \partial\Omega'')}{\mathrm{Tr}(B, \partial\Omega'')}\right) \mathrm{Tr}(B, \partial\Omega'') \quad \mathcal{H}^{d-1}\text{-a.e. on } \partial\Omega'',$$

under the assumption that the components of B and w are bounded and belong to the Sobolev space $W^{1,1}(\Omega'')$. Indeed, the identity is trivial if both w and B are continuous up to the boundary, and the proof of the general case can be immediately achieved by a density argument based on the strong continuity of the trace operator from $W^{1,1}(\Omega'')$ to $L^1(\partial\Omega'', \mathcal{H}^{d-1} \llcorner \partial\Omega'')$ (see for instance Theorem 3.88 of [11]).

Step 2. In this step we prove the general case. Let us apply Gagliardo's theorem (see [33]) on the surjectivity of the trace operator from $W^{1,1}$ into L^1 to obtain a bounded vector field $B_1 \in W^{1,1}(\Omega''; \mathbb{R}^d)$ whose trace on $\partial\Omega' \subset \partial\Omega''$ is equal to the trace of B , seen as a function in $BV(\Omega')$. In particular $\mathrm{Tr}(B, \partial\Omega') = -\mathrm{Tr}(B_1, \partial\Omega'')$. DePining

$$\widehat{B}(x) := \begin{cases} B(x) & \text{if } x \in \Omega', \\ B_1(x) & \text{if } x \in \Omega'', \end{cases}$$

it turns out that $\widehat{B} \in BV(\Omega)$ and that

$$|D\widehat{B}|(\partial\Omega') = 0. \quad (188)$$

Let us consider the function $\theta := \mathrm{Tr}(wB, \partial\Omega') / \mathrm{Tr}(B, \partial\Omega')$ (set equal to 0 wherever the denominator is 0) and let us prove that $\|\theta\|_{L^\infty(\partial\Omega')}$ is less than $\|w\|_{L^\infty(\Omega')}$. Indeed, writing $\partial\Omega'$ as the zeroth-level set of a C^1 function F with $|\nabla F| > 0$ on $\partial\Omega'$ and $\{F = t\} \subset \Omega'$ for $t > 0$ sufficiently small, by Proposition 7.10 we have

$$\begin{aligned} -\|w\|_{L^\infty(\Omega')} \mathrm{Tr}(B, \partial\{F > t\}) &\leq \mathrm{Tr}(wB, \partial\{F > t\}) \\ &\leq \|w\|_{L^\infty(\Omega')} \mathrm{Tr}(B, \partial\{F > t\}) \end{aligned}$$

$\mathcal{H}^{d-1}$ -a.e. on $\{F = t\}$ for $\mathcal{L}^1$ -a.e. $t > 0$ sufficiently small. Passing to the limit as $t \downarrow 0$ and using Theorem 7.11 we recover the same inequality on $\{F = 0\}$, proving the boundedness of θ .

Now, still using Gagliardo's theorem, we can find a bounded function $w_1 \in W^{1,1}(\Omega''; \mathbb{R}^k)$ whose trace on $\partial\Omega'$ is given by θ , so that the normal trace of $w_1 B_1$ on $\partial\Omega''$ is equal to $\mathrm{Tr}(w_1 B, \partial\Omega')$ on the whole of $\partial\Omega'$. DePining

$$\widehat{w}(x) := \begin{cases} w(x) & \text{if } x \in \Omega', \\ w_1(x) & \text{if } x \in \Omega'', \end{cases}$$

by Proposition 7.9 we obtain

$$|D \cdot (\hat{w}^i \hat{B})|(\partial\Omega') = 0, \quad i = 1, \dots, k. \quad (189)$$

Let us apply now Lemma 7.14 and (188), (189), to obtain that the divergence of the vector field $h(\hat{w})\hat{B}$ is a measure with finite total variation in Ω , whose restriction to $\partial\Omega'$ vanishes. As a consequence, Proposition 7.9 gives

$$\text{Tr}^+(h(\hat{w})\hat{B}, \partial\Omega') = \text{Tr}^-(h(\hat{w})\hat{B}, \partial\Omega') \quad \mathcal{H}^{d-1}\text{-a.e. on } \partial\Omega' \quad (190)$$

(here, by a slight abuse of notation, we consider $\partial\Omega'$ as a C^1 oriented surface whose orienting normal coincides with the outer normal to $\partial\Omega'$).

By applying (190), Step 1, and finally, our choice of B_1 and w_1 the following chain of equalities holds $\mathcal{H}^{d-1}$ -a.e. on $\partial\Omega'$:

$$\begin{aligned} \text{Tr}(h(w)B, \partial\Omega') &= \text{Tr}^+(h(\tilde{w})\tilde{B}, \partial\Omega') \\ &= \text{Tr}^-(h(\tilde{w})\tilde{B}, \partial\Omega') \\ &= \text{Tr}(h(w_1)B_1, \partial\Omega'') \\ &= h\left(\frac{\text{Tr}(w_1 B_1, \partial\Omega'')}{\text{Tr}(B_1, \partial\Omega'')}\right) \text{Tr}(B_1, \partial\Omega'') \\ &= h\left(\frac{\text{Tr}(wB, \partial\Omega')}{\text{Tr}(B, \partial\Omega')}\right) \text{Tr}(B, \partial\Omega'). \end{aligned} \quad \square$$

7.4. Proof of Theorem 7.1

In this section we combine the change of variables for traces with a blow-up argument in order to prove Theorem 7.1.

Let Σ be as in the statement. Without loss of generality we can assume that Σ is the boundary of some open set $\Omega' \subset \subset \Omega$, and that the normal ν to Σ is the outer normal of Ω' . Arguing as in the proof of Theorem 7.12, we can build a vector field $\hat{B} \in BV \cap L^\infty(\Omega)$ and a bounded function $\hat{w}$ such that

- $\hat{w} = w$ and $\hat{B} = B$ on $\Omega \setminus \Omega'$;
- $|D \cdot (wB)|(\partial\Omega') = |DB|(\partial\Omega') = 0$.

Given any $x \in \partial\Omega'$, note that

$$\lim_{r \downarrow 0} \frac{|\Omega' \cap B_r^+(x, \nu)|}{r^d} = 0$$

and thus it suffices to prove the claim for $\hat{w}$ and $\hat{B}$. In order to simplify the notation, from now on we will write w and B instead of $\hat{w}$ and $\hat{B}$. Moreover, note that the change of variables for traces implies that $|D \cdot (w^2 B)|(\partial\Omega') = 0$.

Next, fix any $x \in \partial\Omega$ such that $\text{Tr}(B, \partial\Omega')(x) \cdot \nu \neq 0$ and choose a system of coordinates $(x_1, \dots, x_{d-1}, x_d) = (x', x_d)$ in such a way that $\nu = (0, \dots, 0, 1)$.

From now on we simply write $B(x)$ for $\text{Tr}(B, \partial\Omega')(x)$ and for any $r > 0$ consider the $(d-1)$ -dimensional cube

$$C_r := \{x + (y_1, \dots, y_{d-1}, 0) : |y_i| < r\},$$

the d -dimensional parallelogram

$$Q_r := \{y + \rho B(x) : y \in C_r, |\rho| < r\}$$

and the open set $Q_r^+ := Q_r \setminus \bar{\Omega}'$. We denote by 2α the volume of Q_1 (that is, $\alpha = |B(x) \cdot \nu|$).

Clearly, there exists constant C such that $|B_r^+(x, \nu) \setminus Q_r^+| = o(r^d)$, and therefore it suffices to prove that

$$\lim_{r \downarrow 0} \frac{1}{r^d} \int_{Q_r^+(x)} \left| w(y) - \frac{\text{Tr}(wB, \partial\Omega')(x)}{B(x)} \right| dy = 0. \quad (191)$$

We will prove that this holds for any point x which satisfy the following requirements:

(a) x is a Lebesgue point for $\text{Tr}(wB, \partial\Omega')$ and $\text{Tr}(w^2B, \partial\Omega')$, that is,

$$\begin{aligned} \lim_{r \downarrow 0} \frac{1}{r^{d-1}} \int_{\partial\Omega' \cap B_r(x)} & \left[|\text{Tr}(wB, \partial\Omega')(y) - \text{Tr}(wB, \partial\Omega')(x)| \right. \\ & \left. + |\text{Tr}(w^2B, \partial\Omega')(y) - \text{Tr}(w^2B, \partial\Omega')(x)| \right] dy = 0, \end{aligned}$$

and it is a Lebesgue point for B , that is,

$$\lim_{r \downarrow 0} \frac{1}{r^d} \int_{B_r(x)} |B(y) - B(x)| dy = 0;$$

(b) $B(x) \cdot \nu \text{Tr}(w^2B, \partial\Omega') = [\text{Tr}(wB, \partial\Omega')]^2$;

(c) $|D \cdot (wB)|(B_r(x)) + |D \cdot (w^2B)|(B_r(x)) = o(r^{d-1})$.

Since these conditions are satisfied $\mathcal{H}^{d-1}$ -a.e. on the set $\partial\Omega' \setminus \{\text{Tr}(B, \partial\Omega') = 0\}$, this claim will prove the theorem.

Step 1. Let x be any point which satisfies the conditions (a), (b), and (c). In order to simplify the notation, from now on we assume that $x = 0$. Let $r > 0$. Note that using a simple Fubini-type argument we get the existence of an $s(r) \in]r, 2r[$ such that

$$\int_{\partial Q_{s(r)}} |B(y) - B(0)| dy \leq Cr^{-1} \int_{Q_{2r}} |B(y) - B(0)| dy, \quad (192)$$

where C is a constant. Moreover, by Proposition 7.10, we can also assume that, if ζ denotes the outer unit normal to $\partial Q_{s(r)}^+$, then

$$\begin{aligned}\operatorname{Tr}(B, \partial Q_{s(r)}^+) &= B \cdot \zeta \quad \text{and} \\ \operatorname{Tr}(wB, \partial Q_{s(r)}^+) &= wB \cdot \zeta \quad \mathcal{H}^{d-1}\text{-a.e. on } \partial Q_{s(r)}^+.\end{aligned}\tag{193}$$

Denote by B^d the component in direction $(0, \dots, 0, 1) = \nu$ of B and, without loss of generality, assume that $B^d(0) > 0$. Moreover, note that $\alpha = |B(0) \cdot \nu| = B^d(0)$. We will show that

$$\lim_{s \downarrow 0} s(r)^{-d} \int_{Q_{s(r)}^+} w(y) B^d(y) \, dy = \alpha \operatorname{Tr}(wB, \partial \Omega')(0) \tag{194}$$

and

$$\lim_{r \downarrow 0} s(r)^{-d} \int_{Q_{s(r)}^+} w^2(y) B^d(y) \, dy = \alpha (\operatorname{Tr}(wB, \partial \Omega')(0))^2. \tag{195}$$

This will complete the proof, because

$$\begin{aligned}& \lim_{r \downarrow 0} s(r)^{-d} \int_{Q_{s(r)}^+} \left| w(y) - \frac{\operatorname{Tr}(wB, \partial \Omega')(0)}{B(0)} \right| \, dy \\& \leq \lim_{r \downarrow 0} s(r)^{-d} \int_{Q_r^+} |w(y) B^d(0) - \operatorname{Tr}(wB, \partial \Omega')(0)| \, dy \\& = \lim_{r \downarrow 0} s(r)^{-d} \int_{Q_{s(r)}^+} [w^2(y) (B^d(y))^2 - \operatorname{Tr}(wB, \partial \Omega')(0) w(y) B^d(y)] \, dy \\& \quad + [\operatorname{Tr}(wB, \partial \Omega')(0)]^2 \alpha \\& = \lim_{r \downarrow 0} s(r)^{-d} \int_{Q_{s(r)}^+} w^2(y) B^d(y) B^d(0) \, dy - \alpha [\operatorname{Tr}(wB, \partial \Omega')(0)]^2 \\& = \alpha B^d(0) \operatorname{Tr}(w^2 B, \partial \Omega')(0) - \alpha [\operatorname{Tr}(wB, \partial \Omega')(0)]^2 = 0.\end{aligned}$$

Step 2. In this step we show (194). The proof of (195) is completely analogous and therefore we omit it.

Denote by $D_{s(r)}$ the top face of $\partial Q_{s(r)}^+$, that is,

$$D_{s(r)} = \{(y_1, \dots, y_{d-1}, 0) + s(r)B : |y_i| \leq s(r)\}.$$

Then consider the test function $\varphi_r(y) := s(r)B^d(0) - y_d$ and apply the definition of weak trace to get

$$\begin{aligned} - \int_{Q_{s(r)}^+} w(y)B^d(y) \, dy &= - \int_{Q_{s(r)}^+} \varphi_r \, d[D \cdot (wB)] \\ &\quad + \int_{\partial Q_{s(r)}^+ \setminus D_{s(r)}} \varphi_r \operatorname{Tr}(wB, \partial Q_{s(r)}^+) \, d\mathcal{H}^{d-1}. \end{aligned}$$

Recall that for some constant C we have $B_{C^{-1}r}(0) \subset Q_r \subset B_{Cr}(0)$. Therefore the first integral in the right-hand side is $o(s(r)^d)$ by (c). Next, we split the surface $\partial Q_{s(r)}^+ \setminus D_{s(r)}$ into $\partial\Omega' \cap Q_{s(r)}$ and $L := \partial Q_{s(r)}^+ \setminus (D_{s(r)} \cup \partial\Omega')$. Thus

$$\begin{aligned} \lim_{r \downarrow 0} \frac{1}{(s(r))^d} \int_{Q_{s(r)}^+} w(y)B^d(y) \, dy \\ = \lim_{r \downarrow 0} \frac{1}{(s(r))^d} \int_{\partial\Omega' \cap Q_{s(r)}} \varphi_r \operatorname{Tr}(wB, \partial Q_{s(r)}^+) \\ + \lim_{r \downarrow 0} \frac{1}{(s(r))^d} \frac{1}{(s(r))^d} \int_L \varphi_r \operatorname{Tr}(wB, \partial Q_{s(r)}^+). \end{aligned} \quad (196)$$

Note that $\varphi_r = B^d(0)s(r) + o(s(r)) = \alpha s(r) + o(s(r))$ on $Q_{s(r)} \cap \partial\Omega'$. Moreover, note that $\mathcal{H}^{d-1}(\partial\Omega' \cap Q_{s(r)}) = s(r)^{d-1} + o(s(r)^{d-1})$. Thus, from (a) we conclude that

$$\begin{aligned} \lim_{r \downarrow 0} \frac{1}{(s(r))^d} \int_{\partial\Omega' \cap Q_{s(r)}} \varphi_r \operatorname{Tr}(wB, \partial Q_{s(r)}^+) \\ = \lim_{r \downarrow 0} \frac{\alpha}{(s(r))^{d-1}} \int_{\partial\Omega' \cap Q_{s(r)}} \operatorname{Tr}(wB, \partial Q_{s(r)}^+) \\ = \alpha \operatorname{Tr}(wB, \partial\Omega')(0). \end{aligned} \quad (197)$$

Recall that our goal is to show (194). Thus, taking into account (196) and (197), it remains to show that

$$\lim_{r \downarrow 0} \frac{1}{(s(r))^d} \int_L \varphi_r \operatorname{Tr}(wB, \partial Q_{s(r)}^+) = 0. \quad (198)$$

Note that

$$\left| \int_L \varphi_r \operatorname{Tr}(wB, \partial Q_{s(r)}^+) \right| \leq C s(r) \int_L |\operatorname{Tr}(wB, \partial Q_{s(r)}^+)|. \quad (199)$$

Denote by ζ the normal to ∂L and note that $B(0) \cdot \zeta = 0$. Thus

$$\int_L |\operatorname{Tr}(B, \partial Q_{s(r)}^+)| \stackrel{(193)}{=} \int_L |B \cdot \zeta| \leq \int_L |B(y) - B(0)| \stackrel{(192)}{=} o(s(r)^{d-1}). \quad (200)$$

On the other hand, by (193), $|\text{Tr}(wB, \partial Q_{s(r)}^+)| \leq \|w\|_\infty |\text{Tr}(B, \partial Q_{s(r)}^+)|$, and hence (200) and (199) give (198).

7.5. Proof of Theorem 7.4

Given $B \in BV$, the coarea formula and the Alberti's rank-one theorem induce a natural partition of $|D^c B|$ into rectifiable sets of codimension one. In this section we use this property to show Theorem 7.4 from Theorem 7.1.

Let $B^1, \dots, B^d$ be the components of B . Moreover, recall that $\tilde{B}(x)$ denote the approximate limit of B at x whenever it exists.

Note that $|D^c B| \leq \sum_i |D^c B^i|$. Therefore it suffices to prove (180) for $|D^c B^i|$ -a.e. $x \notin E$. According to Alberti's rank-one theorem, there exist Borel functions $\xi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\zeta : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $D^c B = \xi \otimes \zeta |D^c B|$. So it suffices to prove (180) for $|D^c B^i|$ -a.e. $x \in F$, where F is the set of points x where the approximate limit of B exists and $\zeta(x) \cdot \tilde{B}(x) = 0$.

Recall that for $\mathcal{L}^1$ -a.e. t , the set $\Omega_t := \{B^i > t\}$ is a Caccioppoli set and therefore $D\mathbb{1}_{\Omega_t} = \nu_t \mathcal{H}^{d-1} \llcorner \partial^* \Omega_t$, where $\partial^* \Omega_t$ is a rectifiable set and ν_t the approximate exterior unit normal. From the coarea formula for BV functions (see Theorem 2.10), we have

$$\int_{\Omega} \varphi d|DB^i| = \int_{\mathbb{R}} \int_{\partial^* \Omega_t} \varphi d\mathcal{H}^{d-1} dt.$$

Therefore, it suffices to prove (180) for points x in the set

$$F' := \bigcup_{\{t: \Omega_t \text{ is a Caccioppoli set}\}} \partial^* \Omega_t \cap F.$$

Moreover, recall that

$$\int_{\Omega} \Phi \cdot dDB^i = \int_{\mathbb{R}} \int_{\partial^* \Omega_t} \Phi \cdot \nu_t d\mathcal{H}^{d-1} dt.$$

Thus, for $\mathcal{L}^1$ -a.e. t , we have

$$(a) \quad \zeta|_{\partial_t^* \Omega}(x) = \nu_t(x) \text{ for } \mathcal{H}^{d-1}\text{-a.e. } x.$$

Moreover, note that, for $\mathcal{L}^1$ -a.e. t , we have

$$(b) \quad \tilde{B}|_{\partial_t^* \Omega \cap F}(x) = \text{Tr}(B, \Omega_t)|_{\partial_t^* \Omega \cap F}(x) \text{ for } \mathcal{H}^{d-1}\text{-a.e. } x.$$

Therefore, it suffices to prove the claim for every $x \in F'$ which satisfies (a) and (b).

Next, note that if, for $s < t$, Ω_s and Ω_t are both Caccioppoli sets and $x \in \partial^* \Omega_s \cap \partial^* \Omega_t$, then B^1 cannot have approximate limit at x . Therefore, the sets $\partial^* \Omega_t \cap F'$ are all disjoint, and hence the set E of t 's such that $|D \cdot (wB)|(\partial^* \Omega_t \cap F') > 0$ is at most countable. By the coarea formula, we conclude that

$$|D^c B^i| \left(\bigcup_{t \in E} \partial^* \Omega_t \right) = 0.$$

We finally define the set $F'' \subset F'$ of points $x \in \partial^* \Omega_t$ with t and x such that:

- the approximate limit $\tilde{B}(x)$ of B at x exists and $\zeta(x) \cdot \tilde{B}(x) = 0$;
- Ω_t is a Caccioppoli set and $|D \cdot (wB)|(\partial^* \Omega_t \cap F') = 0$;
- $v_t(x) = \zeta(x)$, and hence $v_t(x) \cdot \tilde{B}(x) = 0$;
- $B(x) = \text{Tr}^+(B, \partial^* \Omega_t)(x)$ (where we take v_t as orienting normal for $\partial^* \Omega_t$).

Summarizing what discussed so far, it suffices to prove (180) for $\mathcal{H}^{d-1}$ -a.e. $x \in \partial^* \Omega_t \cap F''$.

So fix a t such that $\partial^* \Omega_t \cap F'' \neq \emptyset$ and let $\{\Sigma_j\}_j$ be a countable family of C^1 surfaces which cover $\mathcal{H}^{d-1}$ -a.e. $\partial^* \Omega_t$. If we denote by v_j the unit normals to Σ_j we have $v_j = v_t$ $\mathcal{H}^{d-1}$ -a.e. on $\Sigma_j \cap \partial^* \Omega_t$. Thus it suffices to show (180) for $\mathcal{H}^{d-1}$ -a.e. $x \in \Sigma_j \cap \partial^* \Omega_t$ such that $v_j(x) \cdot \text{Tr}(B, \Sigma_j)(x) \neq 0$. From Theorem 7.1, for $\mathcal{H}^{d-1}$ -a.e. such x we have

$$\lim_{r \downarrow 0} \frac{1}{r^d} \int_{B_r^+(x, v)} \left| w(y) - \frac{\text{Tr}^+(wB, \Sigma_j)(x)}{\text{Tr}^+(B, \Sigma_j)}(x) \right| dy = 0 \quad (201)$$

and

$$\lim_{r \downarrow 0} \frac{1}{r^d} \int_{B_r^+(x, v)} \left| w(y) - \frac{\text{Tr}^-(wB, \Sigma_j)(x)}{\text{Tr}^-(B, \Sigma_j)}(x) \right| dy = 0. \quad (202)$$

From the definition of F'' , $\text{Tr}^+(B, \Sigma_j)(x) = \text{Tr}^-(B, \Sigma_j)(x) = \text{Tr}^+(B, \partial^* \Omega_t)(x) = \tilde{B}(x)$ for $\mathcal{H}^{d-1}$ -a.e. $x \in \Sigma_j \cap \partial^* \Omega_t$. Moreover, since $|D \cdot (wB)|(\Sigma_j \cap \partial^* \Omega_t) = 0$, from Proposition 7.9 we conclude $\text{Tr}^+(wB, \Sigma_j)(x) = \text{Tr}^-(wB, \Sigma_j)(x)$ for $\mathcal{H}^{d-1}$ -a.e. $x \in \partial^* \Omega_t \cap \Sigma_j$. Therefore (201) and (202) give the desired claim.

8. Bressan's compactness conjecture and the renormalization conjecture for nearly incompressible BV vector fields

In [17] Bressan proposed the following conjectures.

CONJECTURE 8.1 (Bressan's compactness conjecture). Let $b_n: \mathbb{R}_t \times \mathbb{R}_x^m \rightarrow \mathbb{R}^m$ be smooth maps and denote by Φ_n the solution of the ODEs:

$$\begin{cases} \frac{d}{dt} \Phi_n(t, x) = b_n(t, \Phi_n(t, x)), \\ \Phi_n(0, x) = x. \end{cases} \quad (203)$$

Assume that the fluxes Φ_n are nearly incompressible, i.e., that for some constant C we have

$$C^{-1} \leq \det(\nabla_x \Phi_n(t, x)) \leq C, \quad (204)$$

and that $\|b_n\|_\infty + \|\nabla b_n\|_{L^1}$ is uniformly bounded. Then the sequence $\{\Phi_n\}$ is strongly precompact in L^1_{loc} .

This conjecture was advanced in connection with the Keyfitz and Kranzer system, in particular to provide the existence of suitable weak solutions. Though, as shown in Section 5, one can prove well-posedness for this system bypassing it, Conjecture 8.1 is an interesting and challenging question. In this section we will show some recent partial results on it, contained in [10].

First of all, we note that Bressan's compactness conjecture would follow from the following one.

CONJECTURE 8.2 (Renormalization conjecture). Any nearly incompressible bounded BV vector field has the renormalization property of Definition 3.12.

CONJECTURE 8.2 $\implies$ **CONJECTURE 8.1**. Let $\rho_n := (\text{id}, \Phi_n)_\# \mathcal{L}^{m+1}$ be the density generated by the flows Φ_n . From (204) it follows that $C_1 \geq \rho_n \geq C_1^{-1} > 0$ for some constant $C_1 > 0$.

From the BV compactness theorem and the weak* compactness of L^∞ , it suffices to prove Conjecture 8.1 under the additional assumptions that $b_n \rightarrow b$ strongly in L^1_{loc} for some BV vector field b and that $\rho_n \rightharpoonup^* \rho$ in L^∞ for some bounded ρ . Note that

- $\partial_t \rho_n + D_x \cdot (\rho_n b_n)$ converge to $\partial_t \rho + D_x \cdot (\rho b)$ in the sense of distributions, and thus $\partial_t \rho + D_x \cdot (\rho b) = 0$;
- $\rho \geq C_1^{-1}$;
- $\|b\|_\infty < \infty$.

Hence, b is a bounded nearly incompressible vector field, and if Conjecture 8.2 has an affirmative answer, then b has the renormalization property. In this case we can apply Theorem 3.22 to conclude that Φ_n converges strongly in L^1_{loc} to the unique regular Lagrangian flow generated by b . $\square$

The main result of [10] is the following theorem.

THEOREM 8.3. *Let $b \in BV \cap L^\infty(\mathbb{R}^+ \times \mathbb{R}^m, \mathbb{R}^m)$ be a nearly incompressible vector field. Consider the vector field $B \in BV(\mathbb{R}^+ \times \mathbb{R}^m, \mathbb{R} \times \mathbb{R}^m)$ given by $B := (1, b)$ and denote by E its tangential set (see Definition 7.3). If $|D_{t,x}^c \cdot B|(E) = |D_x^c \cdot b|(E) = 0$, then b has the renormalization property.*

More precisely, we will show:

PROPOSITION 8.4. *Let $\Omega \subset \mathbb{R}^d$, $B \in BV \cap L^\infty(\Omega, \mathbb{R}^d)$ and $\rho, w \in L^\infty(\Omega)$ be such that $D \cdot (\rho B) = D \cdot (w \rho B) = 0$ and $\rho \geq c > 0$. Denote by L the set of Lebesgue points of (ρ, w) . Then for every $h \in C^1(\mathbb{R})$, the measure $D \cdot (\rho h(w) B)$ satisfies the bound $|D \cdot (\rho h(w) B)| \leq C |D^c \cdot B| \llcorner (\Omega \setminus L)$ for some constant C .*

Using the same arguments as in the proof of Lemma 5.10, Theorem 8.3 follows from Proposition 8.4 and Theorem 7.4.

These results naturally raise the following problem:

QUESTION 8.5 (Divergence problem). Let $B \in BV_{\text{loc}} \cap L^\infty_{\text{loc}}(\Omega, \mathbb{R}^d)$. Under which conditions the Cantor part of the divergence $|D^c \cdot B|$ vanishes on the tangential set of B ?

In Section 9 we will prove that indeed some condition is needed, namely we show a planar BV vector field B such that $|D^c \cdot B|$ does not vanish on the tangential set of B . However we do not know the answer to the following question. Note that in view of Theorem 8.3 a positive answer would imply the renormalization conjecture.

QUESTION 8.6. Let $B \in BV_{\text{loc}} \cap L_{\text{loc}}^\infty(\Omega, \mathbb{R}^d)$ and let $\rho \in L^\infty(\Omega)$ be such that $\rho \geq C > 0$ and $D \cdot (\rho B) = 0$. Is it true that $|D^c \cdot B|$ vanishes on the tangential set of B ?

8.1. Absolutely continuous and jump parts of the measure $D \cdot (\rho h(w)B)$

Let B, ρ and w be as in Proposition 8.4. Let c be such that $\rho \geq c$ and define $H : [c, \infty[\times \mathbb{R}$ by $H(r, u) := rh(u/r)$. Clearly H is C^1 and we can extend it to a C^1 function of $\mathbb{R}^2$. Next set $v := \rho w$. Then we have

$$D \cdot (\rho B) = 0 \quad D \cdot (v B) = 0 \quad D \cdot (\rho h(w)B) = D \cdot (H(\rho, v)B)$$

and we can apply Theorem 4.1 in order to get

$$\left| D \cdot (H(\rho, v)B) - \left(H(\rho, v) - \frac{\partial H}{\partial r}(\rho, v)\rho - \frac{\partial H}{\partial u}(\rho, v)v \right) D^a \cdot B \right| \leq C |D^s \cdot B|.$$

On the other hand, since the essential range of (ρ, v) is in $[c, \infty[\times \mathbb{R}$, one immediately sees that

$$H(\rho, v) - \frac{\partial H}{\partial r}(\rho, v)\rho - \frac{\partial H}{\partial u}(\rho, v)v = 0.$$

Hence, we have concluded the following corollary.

COROLLARY 8.7. Let B, ρ, w and h be as in Proposition 8.4. Then $D \cdot (\rho h(w)B)$ is a Radon measure and there exists a constant C such that $|D \cdot (\rho h(w)B)| \leq C |D^s \cdot B|$.

We will next use the trace properties of divergence measure fields in order to show the following proposition.

PROPOSITION 8.8. Let B, ρ, w and h be as in Proposition 8.4. Then there exists a constant C such that $|D \cdot (\rho h(w)B)| \leq C |D^c \cdot B|$.

PROOF. Consider the jump set J_B of B , its approximate unit normal v and the approximate left and right traces of B on J_B . Then $|D^j \cdot B| = |(B^+ - B^-) \cdot v| \mathcal{H}^{d-1} \llcorner J_B$ and, by Corollary 8.7,

$$|D \cdot (\rho h(w)B)| \leq C |(B^+ - B^-) \cdot v| \mathcal{H}^{d-1} \llcorner J_B + C |D^c \cdot B|. \quad (205)$$

Now, let $\{\Sigma_i\}_i$ be a countable family of hypersurfaces such that $B \subset \bigcup_i \Sigma_i$. In order to complete the proof it suffices to show that $D \cdot (\rho h(w)B) \llcorner \Sigma_i = 0$ for every i . Next, fix any $\varepsilon > 0$ such that $\varepsilon \leq \rho$ a.e. and consider the function $F_\varepsilon : (-\infty, -\varepsilon] \cup [\varepsilon, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ defined by $F_\varepsilon(r, u) := h(u/r)$. Extend it to a C^1 function defined on all $\mathbb{R}^2$. Next set $H_\varepsilon(r, u) := rh(u/r)$. Then, recalling that $D \cdot (\rho B) = 0$ and $D \cdot ((\rho w)B) = 0$, we can use Proposition 7.9 and Theorem 7.12 to get

$$\begin{aligned} & [D \cdot (h(w)\rho B)] \llcorner \Sigma \\ &= [D \cdot (H_\varepsilon(w\rho, \rho))B] \llcorner \Sigma \\ &= \left[H_\varepsilon \left(\frac{\text{Tr}^+(w\rho B, \Sigma)}{\text{Tr}^+(B, \Sigma)}, \frac{\text{Tr}^+(\rho B, \Sigma)}{\text{Tr}^+(B, \Sigma)} \right) \text{Tr}^+(B, \Sigma) \right. \\ & \quad \left. - H_\varepsilon \left(\frac{\text{Tr}^-(w\rho B, \Sigma)}{\text{Tr}^-(B, \Sigma)}, \frac{\text{Tr}^-(\rho B, \Sigma)}{\text{Tr}^-(B, \Sigma)} \right) \text{Tr}^-(B, \Sigma) \right] \mathcal{H}^{d-1} \llcorner \Sigma. \end{aligned} \quad (206)$$

Now consider the set

$$\Sigma' := \{x \in \Sigma : \text{Tr}^+(B, \Sigma)(x) = 0 \text{ or } \text{Tr}^-(B, \Sigma)(x) = 0\}.$$

Applying Theorem 7.12 to $H \equiv 1$, we conclude that, up to $\mathcal{H}^{d-1}$ -negligible sets,

$$\Sigma' \subset \Sigma_0 := \{x \in \Sigma : \text{Tr}^-(\rho B, \Sigma)(x) = 0 \text{ or } \text{Tr}^+(\rho B, \Sigma)(x) = 0\}.$$

Next note that, by Proposition 7.9,

$$0 = D \cdot (\rho B) \llcorner \Sigma = [\text{Tr}^+(\rho B, \Sigma) - \text{Tr}^-(\rho B, \Sigma)] \mathcal{H}^{d-1} \llcorner \Sigma.$$

and

$$0 = D \cdot (\rho w B) \llcorner \Sigma = [\text{Tr}^+(\rho w B, \Sigma) - \text{Tr}^-(\rho w B, \Sigma)] \mathcal{H}^{d-1} \llcorner \Sigma.$$

Thus, we conclude that $\text{Tr}^-(\rho B, \Sigma) = \text{Tr}^+(\rho B, \Sigma)$ and $\text{Tr}^+(\rho w B, \Sigma) = \text{Tr}^-(\rho w B, \Sigma)$ a.e. on Σ . Recall the definition of H_ε . Then

- the expression

$$\begin{aligned} E := & H_\varepsilon \left(\frac{\text{Tr}^+(w\rho B, \Sigma)}{\text{Tr}^+(B, \Sigma)}, \frac{\text{Tr}^+(\rho B, \Sigma)}{\text{Tr}^+(B, \Sigma)} \right) \text{Tr}^+(B, \Sigma) \\ & - H_\varepsilon \left(\frac{\text{Tr}^-(w\rho B, \Sigma)}{\text{Tr}^-(B, \Sigma)}, \frac{\text{Tr}^-(\rho B, \Sigma)}{\text{Tr}^-(B, \Sigma)} \right) \text{Tr}^-(B, \Sigma) \end{aligned}$$

vanishes $\mathcal{H}^{d-1}$ -a.e. on Σ_0 .

- $\mathcal{H}^{d-1}$ -a.e. on $\Sigma_\varepsilon := \{|\operatorname{Tr}^+(\rho B, \Sigma)| \geq \varepsilon\}$ we have $|\operatorname{Tr}^-(\rho B, \Sigma)| \geq \varepsilon$ and $\operatorname{tr}^+(B, \Sigma) \neq 0 \neq \operatorname{Tr}^-(B, \Sigma)$. Thus we can compute

$$E = h\left(\frac{\operatorname{Tr}^+(\rho w B, \Sigma)}{\operatorname{Tr}^-(\rho B, \Sigma)}\right) \operatorname{Tr}^+(\rho B, \Sigma) - h\left(\frac{\operatorname{Tr}^-(\rho w B, \Sigma)}{\operatorname{Tr}^-(\rho B, \Sigma)}\right) \operatorname{Tr}^-(\rho B, \Sigma).$$

Recalling that $\operatorname{Tr}^+(\rho B, \Sigma) = \operatorname{Tr}^-(\rho B, \Sigma)$ and $\operatorname{Tr}^+(\rho w B, \Sigma) = \operatorname{Tr}^-(\rho w B, \Sigma)$, we conclude that E vanishes $\mathcal{H}^{d-1}$ -a.e. on Σ_ε .

Therefore, by (206) we have

$$0 = [D \cdot (\rho h(w) B)] \llcorner \{x \in \Sigma: 0 < |\operatorname{Tr}^+(\rho B, \Sigma)(x)| < \varepsilon\}.$$

Letting $\varepsilon \downarrow 0$ we get $D \cdot (\rho h(w) B) \llcorner \Sigma = 0$, which is the desired conclusion. $\square$

8.2. Proof of Proposition 8.4 and concentration of commutators

In the previous section we proved that, under the assumptions of Proposition 8.4, $|D \cdot (\rho h(w) B)| \leq C |D^c \cdot B|$. Here we will state a new commutator estimate and with the help of it we will complete the proof of Proposition 8.4.

As in the previous section,

- we fix w, ρ, b and h as in Proposition 8.4;
- we let $c > 0$ be such that $c < \rho$ a.e. and we define $H: [c, \infty[\times \mathbb{R} \rightarrow \mathbb{R}$ setting $H(r, u) := rh(u/r)$;
- we extend H to a C^1 function on $\mathbb{R}^2$.

Next we fix a nonnegative kernel $\eta \in C_c^\infty(\mathbb{R}^d)$ and consider the standard family of mollifiers $\{\eta_\varepsilon\}_{\varepsilon>0}$. If we set $v := \rho w$, then $D \cdot (\rho h(w) B) = D \cdot (H(\rho, v) B)$ is the weak limit of

$$\begin{aligned} & D \cdot (H(\rho * \eta_\varepsilon, v * \eta_\varepsilon) B) \\ &= \left[\frac{\partial H}{\partial r}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) D(\rho * \eta_\varepsilon) \cdot B + \frac{\partial H}{\partial u}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) D(v * \eta_\varepsilon) \cdot B \right] \\ &\quad + H(\rho * \eta_\varepsilon, v * \eta_\varepsilon) D \cdot B \\ &= \left[\frac{\partial H}{\partial r}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) D \cdot (\rho * \eta_\varepsilon B) + \frac{\partial H}{\partial u}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) D \cdot (v * \eta_\varepsilon B) \right] \\ &\quad + \left[H(\rho * \eta_\varepsilon, v * \eta_\varepsilon) - \frac{\partial H}{\partial r}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) \rho * \eta_\varepsilon \right. \\ &\quad \left. + \frac{\partial H}{\partial u}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) v * \eta_\varepsilon \right] D \cdot B. \end{aligned}$$

Next, note that the range of $\rho * \eta_\varepsilon$ is contained in $[c, \infty[$. Thus, from the definition of H it follows that it is a 1-homogeneous function on the range of $(\rho * \eta_\varepsilon, v * \varepsilon)$. This implies that

$$\frac{\partial H}{\partial r}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) \rho * \eta_\varepsilon + \frac{\partial H}{\partial u}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) v * \eta_\varepsilon - H(\rho * \eta_\varepsilon, v * \eta_\varepsilon) = 0.$$

Recalling that $D \cdot ((\rho B) * \eta_\varepsilon) = D \cdot ((v B) * \eta_\varepsilon) = 0$ we conclude that $D \cdot (H(\rho, v)B)$ is the limit, in the distributional sense, of the expressions

$$\begin{aligned} & \frac{\partial H}{\partial r}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) [D \cdot (\rho * \eta_\varepsilon B) - D \cdot (\rho B) * \eta_\varepsilon] \\ & + \frac{\partial H}{\partial u}(\rho * \eta_\varepsilon, v * \eta_\varepsilon) [D \cdot (v * \eta_\varepsilon B) - D \cdot (v B) * \eta_\varepsilon]. \end{aligned} \quad (207)$$

This discussion justifies the introduction of the following notation and terminology.

DEFINITION 8.9. Let $\Omega \subset \mathbb{R}^d$, $B \in BV(\Omega, \mathbb{R}^d)$, $z \in L^\infty(\Omega, \mathbb{R}^k)$ and $H \in C^1(\mathbb{R}^k)$. If $\{\eta_\varepsilon\}_{\varepsilon>0}$ is a standard family of mollifiers, then we define the commutators

$$\begin{aligned} T_\delta^i &:= (D \cdot (z^i B)) * \eta_\delta - D \cdot (z^i * \eta_\delta B) \\ T_\delta^i &:= \frac{\partial H}{\partial u_i}(z * \eta_\delta) T_\delta^i. \end{aligned}$$

Note that the commutators T_δ^i coincide with the commutators T_δ^i of Definition 4.2. Recalling Proposition 4.6, we conclude that the distributions T_δ^i are measures with uniformly bounded total variations. Then Proposition 8.4 follows from the following theorem, which will be proved in the next section.

THEOREM 8.10 (Commutator estimate). Let T_δ^i be as in Definition 8.9 and consider the set L_z of Lebesgue points of z . Then any weak* limit of T_δ^i is a measure ν such that $|\nu|(\Omega \setminus L_z) = 0$.

8.3. Proof of Theorem 8.10

Recalling the proof of Proposition 4.6, T_δ^i can be written as $r_\delta^i \mathcal{L}^d - (z^i * \eta_\delta) D \cdot B$, where

$$r_\delta^i(x) := \int_{\mathbb{R}^d} z^i(x') [(B(x) - B(x')) \cdot \nabla \eta_\delta(x' - x)] dx'. \quad (208)$$

An important step toward the proof of Theorem 8.10 is the following representation lemma.

LEMMA 8.11 (Double averages lemma). *Let $\Phi \in L^\infty(\Omega)$ and assume that its support is a compact subset of Ω . Then, for δ sufficiently small, we have*

$$\int_{\mathbb{R}^d} \Phi(x) r_\delta^i(x) dx = \sum_{j,l} \int_{\mathbb{R}^d} A_\delta^{ijl}(\xi) d[D_l B^j](\xi), \quad (209)$$

where the functions A_δ^{ijl} are given by the double average

$$A_\delta^{ijl}(\xi) := -\frac{1}{\delta} \int_0^\delta \int_{\mathbb{R}^d} y_l \frac{\partial \eta}{\partial x_j}(y) \Phi(\xi - \tau y) z^i(\xi + (\delta - \tau)y) dy d\tau. \quad (210)$$

PROOF. Fix $\Phi \in L^\infty(\Omega)$ and with compact support contained in Ω . Then, if δ is sufficiently small, A_δ^{ijl} has compact support contained in Ω . We now prove that A_δ^{ijl} is a continuous function. Taking into account that Φ and z are bounded, it suffices to show that

$$R_\varepsilon(\xi) := \int_\varepsilon^{\delta-\varepsilon} \int_{\mathbb{R}^d} y_l \frac{\partial \eta}{\partial x_j}(y) \Phi(\xi - \tau y) z^i(\xi + (\delta - \tau)y) dy d\tau$$

is continuous for any $\varepsilon \in]0, \delta/2[$. This claim can be proved as follows. First of all, without loss of generality, we can assume that both z and Φ are compactly supported. Next we take sequences $\{z_n\}$ and $\{\Phi_n\}$ of continuous compactly supported functions such that $\|z - z_n\|_{L^2} + \|\Phi - \Phi_n\|_{L^2} \downarrow 0$. If we set

$$R_{n,\varepsilon}(\xi) := \int_\varepsilon^{\delta-\varepsilon} \int_{\mathbb{R}^d} y_l \frac{\partial \eta}{\partial x_j}(y) \Phi_n^l(\xi - \tau y) z_n^i(\xi + (\delta - \tau)y) dy d\tau,$$

then each $R_{n,\varepsilon}$ is continuous. Moreover one can easily check that

$$|R_{n,\varepsilon}(\xi) - R_\varepsilon(\xi)| \leq C\delta\varepsilon^{-d} (\|\Phi\|_{L^2} \|z - z_n\|_{L^2} + \|z_n\|_{L^2} \|\Phi_n - \Phi\|_{L^2}).$$

Therefore $R_{n,\varepsilon} \rightarrow R_\varepsilon$ uniformly, and we conclude that R_ε is continuous.

Now, fix B and δ as in the statement of the lemma. We approximate B in L^1_{loc} with a sequence of smooth functions B_n , in such a way that $D_k B_n^j$ converge weakly* to $D_k B^j$ on Ω . Hence, we have that

$$R_n^i(x) := \int_{\mathbb{R}^d} z^i(x') [(B_n(x) - B_n(x')) \cdot \nabla \eta_\delta(x' - x)] dx'$$

converge strongly in L^1_{loc} to r_δ^i . Moreover, since A_δ^{ijl} is a continuous and compactly supported function, we have

$$\lim_{n \rightarrow \infty} \int A_\delta^{ijl}(\xi) d[D_l B_n^j](\xi) = \int A_\delta^{ijl}(\xi) d[D_l B^j](\xi).$$

Hence it is enough to prove the statement of the lemma for B_n , which are smooth functions.

Thus, we fix a smooth function B and compute

$$\begin{aligned}
 & - \int r_\delta^i(x) \Phi(x) dx \\
 &= - \int_{\mathbb{R}^d} \Phi(x) \int_{\mathbb{R}^d} z^i(x') [(B(x) - B(x')) \cdot \nabla \eta_\delta(x' - x)] dx' dx \\
 &= - \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi(x) z^i(x + \delta y) \frac{B(x) - B(x + \delta y)}{\delta} \cdot \nabla \eta(y) dy dx \\
 &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi(x) z^i(x + \delta y) \frac{1}{\delta} \int_0^\delta \sum_{l,j} y_l \frac{\partial B^j}{\partial x_l}(x + \tau y) \frac{\partial \eta}{\partial x_j}(y) d\tau dy dx \\
 &= \sum_{k,l} \int_{\mathbb{R}^d} \left[\frac{1}{\delta} \int_0^\delta \int_{\mathbb{R}^d} y_l \frac{\partial \eta}{\partial x_j}(y) \Phi(\xi - \tau y) z^i(\xi + (\delta - \tau)y) dy d\tau \right] \frac{\partial B^j}{\partial x_l}(\xi) d\xi.
 \end{aligned}$$

Since the measure $\frac{\partial B^j}{\partial x_l} \mathcal{L}^d$ is equal to $D_l B^j$, the claim of the lemma follows. $\square$

PROOF OF THEOREM 8.10. We rewrite $\mathcal{T}_\delta^i$ as

$$\mathcal{T}_\delta^i = \frac{\partial H}{\partial u_i}(z * \eta_\delta) r_\delta^i \mathcal{L}^d - \frac{\partial H}{\partial u_i}(z * \eta_\delta) (z^i * \eta_\delta) D \cdot B. \quad (211)$$

We define the matrix-valued measures

$$\begin{aligned}
 \alpha &:= DB \llcorner L_z, \\
 \beta &:= DB \llcorner (\Omega \setminus L_z)
 \end{aligned}$$

and the measures

$$\begin{aligned}
 \gamma &:= [D \cdot B] \llcorner L_z, \\
 \lambda &:= [D \cdot B] \llcorner (\Omega \setminus L_z).
 \end{aligned}$$

Then we introduce the measures S_δ^i and R_δ^i given by the following linear functionals on $\varphi \in C_c(\Omega)$:

$$\begin{aligned}
 \langle S_\delta^i, \varphi \rangle &:= \sum_{j,l} \int_{\mathbb{R}^d} g_\delta^{ijl}(\xi) d[\alpha_{lj}](\xi) \\
 &\quad - \int_{\mathbb{R}^d} \varphi(x) \frac{\partial H}{\partial u_i}(z * \eta_\delta(x)) z^i * \eta_\delta(x) d\gamma(x),
 \end{aligned} \quad (212)$$

and

$$\begin{aligned} \langle R_\delta^i, \varphi \rangle &:= \sum_{j,l} \int_{\mathbb{R}^d} g_\delta^{ijl}(\xi) d[\beta_{lj}](\xi) \\ &\quad - \int_{\mathbb{R}^d} \varphi(x) \frac{\partial H}{\partial u_i}(z * \eta_\delta(x)) z^i * \eta_\delta(x) d\lambda(x), \end{aligned} \quad (213)$$

where

$$\begin{aligned} g_\delta^{ijl}(\xi) &:= -\frac{1}{\delta} \int_0^\delta \int_{\mathbb{R}^d} y_l \frac{\partial \eta}{\partial x_j}(y) \varphi(\xi - \tau y) \\ &\quad \times \frac{\partial h}{\partial u_i}(z * \eta_\delta(\xi - \tau y)) z^i (\xi + (\delta - \tau)y) dy d\tau. \end{aligned} \quad (214)$$

This formula for g_δ^{ijl} comes from the formulas for A_δ^{ijl} of Lemma 8.11, where we choose as Φ the function

$$\Phi := \varphi \frac{\partial H}{\partial u_i}(z * \eta_\delta).$$

Hence, comparing (214) with (211) and (210), from Lemma 8.11 we conclude that $T_\delta^i = S_\delta^i + R_\delta^i$.

Let R_0^i be any weak limit of a subsequence $\{R_{\delta_n}^i\}_{\delta_n \downarrow 0}$ and let S_0^i be any weak limit of a subsequence (not relabeled) of $\{S_{\delta_n}^i\}$. In what follows we will prove that

- (i) $R_0^i \ll |\lambda| + |\beta|$,
- (ii) $S_0^i = 0$.

Since $|\lambda|$ and $|\beta|$ are concentrated on $\Omega \setminus L_z$, (i) and (ii) prove the theorem.

Proof of (i). Let us fix a smooth function φ with $|\varphi| \leq 1$ and with support $K \subset \subset \Omega$. If we define g_δ^{ijl} as in (214), there exists a constant C , depending only on w and H , such that $\|g_\delta^{ijl}\|_\infty \leq C$. Hence, it follows that

$$\left| \int \varphi dR_\delta^i \right| \leq C \|z\|_\infty \left\{ |\beta| \left(\bigcup_{j,l} \text{supp}(g_\delta^{ijl}) \right) + |\lambda|(K) \right\}. \quad (215)$$

Moreover, it is easy to check that, if K_ε denotes the ε -neighborhood of K , then $\text{supp}(g_\delta^{ijl}) \subset K_{2\delta}$. Hence, passing into the limit in (215), we conclude that

$$\left| \int \varphi dR_0^i \right| \leq C \|z\|_\infty (|\lambda|(K) + |\beta|(K)).$$

From the arbitrariness of $\varphi \in C_c^\infty(\tilde{\Omega})$ it follows easily that $R_0^i \ll \tilde{\Omega} \leq C(|\beta| + |\lambda|)$.

Proof of (ii). By definition of L_z , z has Lebesgue limit $\tilde{z}(x)$ at every $x \in L_z$. Hence it follows that

$$\lim_{\delta \downarrow 0} z * \eta_\delta(x) = \tilde{z}(x). \quad (216)$$

Fix φ and define g_δ^{ijl} as in (214). We will show that, for every $\xi \in L_z$, we have that

$$\lim_{\delta \downarrow 0} g_\delta^{ijl}(\xi) = g^{ijl}(\xi), \quad (217)$$

where

$$g^{ijl}(\xi) := -\varphi(\xi) \frac{\partial H}{\partial u_i}(\tilde{z}(\xi)) \tilde{z}^i(\xi) \int_{\mathbb{R}^d} y_l \frac{\partial \eta}{\partial x_j}(y) dy.$$

Integrating by parts we get

$$g^{ill}(\xi) = \varphi(\xi) \frac{\partial H}{\partial u_i}(\tilde{z}(\xi)) \tilde{z}^i(\xi), \quad (218)$$

$$P g^{ijl}(\xi) = 0 \quad \text{for } j \neq l. \quad (219)$$

Recall that g_δ^{ijl} , φ , $z * \eta_\delta$, $H(z * \eta_\delta)$ and $\nabla H(z * \eta_\delta)$ are all uniformly bounded. Hence, letting $\delta \downarrow 0$ in (212), from (216)–(219), and the dominated convergence theorem we conclude that

$$\begin{aligned} \langle S_0^i, \varphi \rangle &= \sum_l \int_{\mathbb{R}^d} \frac{\partial H}{\partial u_i}(\tilde{z}(\xi)) \tilde{z}^i(\xi) \varphi(\xi) d[\alpha_{ll}](\xi) \\ &\quad - \int_{\mathbb{R}^d} \frac{\partial H}{\partial u_i}(\tilde{z}(x)) \tilde{z}^i(x) \varphi(x) d\gamma(x). \end{aligned}$$

Recalling that $\sum_l \alpha_{ll} = \sum_l D_l^c B^l \llcorner L_z = D^c \cdot B \llcorner L_z$ and $\gamma = D^c \cdot B \llcorner L_z$, we conclude that $\langle S_0^i, \varphi \rangle = 0$. The arbitrariness of φ gives (ii).

Hence, to finish the proof, it suffices to show (217). Recalling the smoothness of φ and the fact that η is supported in the ball $B_1(0)$ we conclude that it suffices to show that

$$\begin{aligned} I_\delta := \frac{1}{\delta} \int_0^\delta \int_{B_1(0)} &\left| \frac{\partial H}{\partial u_j}(z * \eta_\delta(\xi - \tau y)) z^i(\xi + (\delta - \tau)y) \right. \\ &\left. - \frac{\partial H}{\partial u_j}(\tilde{z}(\xi)) \tilde{z}^i(\xi) \right| dy d\tau \end{aligned} \quad (220)$$

converges to 0. Then, we write

$$\begin{aligned} I_\delta &\leq \frac{1}{\delta} \int_0^\delta \int_{B_1(0)} \left| \frac{\partial H}{\partial u_j}(z * \eta_\delta(\xi - \tau y)) - \frac{\partial h}{\partial u_j}(\tilde{z}(\xi)) \right| |z^i(\xi + (\delta - \tau)y)| dy d\tau \\ &\quad + \frac{1}{\delta} \int_0^\delta \int_{B_1(0)} \left| \frac{\partial H}{\partial u_j}(\tilde{z}(\xi)) \right| |z^i(\xi + (\delta - \tau)y) - \tilde{z}^i(\xi)| dy d\tau \end{aligned}$$

$$\begin{aligned}
&\leq \frac{C_1}{\delta} \int_0^\delta \int_{B_1(0)} |z * \eta_\delta(\xi - \tau y) - \tilde{z}(\xi)| \, dy \, d\tau \\
&\quad + \frac{C_2}{\delta} \int_0^\delta \int_{B_1(0)} |z(\xi + (\delta - \tau)y) - \tilde{z}(\xi)| \, d\xi \, d\tau \\
&=: C_1 J_\delta^1 + C_2 J_\delta^2,
\end{aligned}$$

where the constants C_1 and C_2 depend only on ξ , z and H . Note that

$$\begin{aligned}
J_\delta^1 &= \frac{1}{\delta} \int_0^\delta \int_{B_1(0)} |z(\xi + \tau y) - \tilde{z}(\xi)| \, dy \, d\tau \\
&= \frac{1}{\delta} \int_0^\delta \left[\frac{1}{\tau^d} \int_{B_\tau(\xi)} |z(y') - \tilde{w}(\xi)| \, dy' \right] d\tau
\end{aligned}$$

and

$$\begin{aligned}
J_\delta^2 &= \frac{1}{\delta} \int_0^\delta \int_{B_1(0)} |z * \eta_\delta(\xi + \tau y) - \tilde{z}(\xi)| \, dy \, d\tau \\
&= \frac{1}{\delta} \int_0^\delta \left[\frac{1}{\tau^d} \int_{B_\tau(\xi)} |z * \eta_\delta(y') - \tilde{z}(\xi)| \, dy' \right] d\tau.
\end{aligned}$$

Hence, since $\tilde{z}(\xi)$ is the Lebesgue limit of z at ξ , we conclude that $J_\delta^1 + J_\delta^2 \rightarrow 0$. This completes the proof. $\square$

9. Tangential sets of BV vector fields

In this section we will show the following proposition.

PROPOSITION 9.1. *There exists $B \in BV \cap L^\infty(\mathbb{R}^2, \mathbb{R}^2)$ such that $|D^c \cdot B|(E) > 0$, where E denotes the tangential set of B .*

As already explained in Section 8, this proposition motivates Question 8.5 and in particular Question 8.6. There are other natural conditions under which it would be interesting to investigate the validity of $|D^c \cdot B|(E) = 0$, such as

- $B = \nabla \alpha \in BV_{\text{loc}}(\Omega)$ for some $\alpha \in W_{\text{loc}}^{1,\infty}$ (in this case $D \cdot B = \Delta \alpha$);
- B is a (semi)-monotone operator, that is

$$\langle B(y) - B(x), y - x \rangle \geq \lambda |x - y|^2 \quad \forall x, y \in \Omega. \quad (221)$$

- B is both curl-free and (semi)-monotone.

PROOF OF PROPOSITION 9.1. We set $\Omega := \{(x, y) \in \mathbb{R}^2: 1 < x < 2, 0 < y < x\}$. We construct a scalar function $u \in L^\infty \cap BV(\Omega)$ with the following properties:

(a) $D_y^c u \neq 0$;

(b) $D_x u + D_y(u^2/2)$ is a pure jump measure, i.e., it is concentrated on the jump set J_u . Given such a function u , the field $B = (1, u)\mathbb{1}_\Omega$ meets the requirements of the proposition. Indeed, let $\tilde{B} = (1, \tilde{u})\mathbb{1}_\Omega$ be the precise representative of B . Due to (b) the Cantor part of $D_x u + D_y(u^2/2)$ vanishes. Hence using the chain rule of Vol'pert we get

$$D_x^c u + \tilde{u} D_y^c u = 0. \quad (222)$$

Denote by $M(x)$ the Radon-Nikodym derivative $DB/|DB|$. Then we have

$$\begin{aligned} M \cdot \tilde{B} |D^c B| &= D^c B \cdot \tilde{B} \\ &= \begin{pmatrix} 0 & 0 \\ D_x^c u & D_y^c u \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} 0 \\ D_x^c u + \tilde{u} D_y^c u \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Hence we conclude that $M(x) \cdot \tilde{B}(x) = 0$ for $|D^c B|$ -a.e. x , that is, $|D^c B|$ is concentrated on the tangential set E of B . Therefore $|D^c \cdot B|(\Omega \setminus E) = 0$. On the other hand, from (a) we have $D^c \cdot B = D_y^c u \neq 0$. Hence we conclude $|D^c \cdot B|(E) > 0$.

We now come to the construction of the desired u . This is achieved as the limit of a suitable sequence of functions u_k .

Step 1: Construction of u_k . Consider the auxiliary 1-periodic function $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$\sigma(p+x) = 1-x, \quad 0 < x \leq 1, \quad p \in \mathbb{Z}.$$

We let $\gamma_k : [0, 1] \rightarrow [0, 1]$ be the usual piecewise linear approximation of the Cantor ternary function, that is $\gamma_0(z) = z$ and, for $k \geq 1$,

$$\gamma_k(z) = \begin{cases} \frac{1}{2} \gamma_{k-1}(3z), & 0 < z \leq \frac{1}{3}, \\ \frac{1}{2}, & \frac{1}{3} < z \leq \frac{2}{3}, \\ \frac{1}{2} (1 + \gamma_{k-1}(3z-2)), & \frac{2}{3} < z \leq 1. \end{cases}$$

Notice that

$$\gamma'_k(z) \in \left\{ 0, \left(\frac{3}{2}\right)^k \right\} \quad (223)$$

and

$$|\gamma_k(z) - \gamma_{k-1}(z)| \leq \frac{1}{3} \cdot 2^{-k}. \quad (224)$$

We set $G :=]1, 2[\times]0, 1[$ and we define $\varphi_k : G \rightarrow \mathbb{R}$ by

$$\varphi_k(x, z) = xz + \sum_{j=1}^k 4^{1-j} \sigma(4^{j-1}x) (\gamma_{j-1}(z) - \gamma_j(z)).$$

Note that φ_k is bounded. To describe more precisely the behavior of this function we introduce the following sets: The strips

$$S_i^k :=]1 + (i-1)4^{1-k}, 1 + i4^{1-k}[\times \mathbb{R}, \quad i = 1, \dots, 4^{k-1},$$

and the vertical lines

$$V_i^k := \{i4^{1-k}\} \times \mathbb{R}, \quad i = 1, \dots, 4^{k-1} - 1.$$

Then φ_k is Lipschitz on each rectangle $S_i^k \cap G$ and it has jump discontinuities on the segments $V_i^k \cap G$. Therefore φ_k is a BV function and satisfies the identities $D_x \varphi_k = D_x^j \varphi_k + D_x^a \varphi_k$ and $D_y \varphi_k = D_y^a \varphi_k$. Moreover, denoting by $(\partial_x \varphi_k, \partial_y \varphi_k)$ the density of the absolutely continuous part of the derivative, we get

$$\begin{aligned} \partial_x \varphi_k(x, z) &= z + (\gamma_1(z) - z) + (\gamma_2(z) - \gamma_1(z)) + \dots + (\gamma_k(z) - \gamma_{k-1}(z)) \\ &= \gamma_k(z). \end{aligned} \quad (225)$$

Clearly

$$0 \leq 4^{1-j} \sigma(4^{j-1}x) - 4^{-j} \sigma(4^j x) \leq 3 \cdot 4^{-j}.$$

Therefore, using also (223), on each rectangle $S_i^k \cap G$ we can estimate

$$\begin{aligned} \partial_z \varphi_k(x, z) &= x + \sigma(x) - (\sigma(x) - 4^{-1} \sigma(4x)) \gamma_1'(z) \\ &\quad - (4^{-1} \sigma(4x) - 4^{-2} \sigma(4^2 x)) \gamma_2'(z) - \dots \\ &\quad - (4^{2-k} \sigma(4^{k-1} x) - 4^{1-k} \sigma(4^{k-1} x)) \gamma_{k-1}'(z) \\ &\quad - 4^{1-k} \sigma(4^{k-1} x) \gamma_k'(z) \\ &\geq 2 - 3(4^{-1} \gamma_1'(z) + \dots + 4^{1-k} \gamma_k'(z)) - 4^{1-k} \gamma_k'(z) \\ &\geq 2 - 3 \left(\frac{3}{8} + \dots + \left(\frac{3}{8} \right)^{k-1} \right) - 4 \left(\frac{3}{8} \right)^k. \end{aligned}$$

Since

$$4 \left(\frac{3}{8} \right)^k \leq 3 \left(\left(\frac{3}{8} \right)^k + \left(\frac{3}{8} \right)^{k+1} + \dots \right),$$

we obtain

$$\partial_z \varphi_k \geq 2 - 3 \left(\frac{3}{8} + \left(\frac{3}{8} \right)^2 + \dots \right) = \frac{1}{5}. \quad (226)$$

Hence, since $\varphi_k(x, \cdot)$ maps $[0, 1]$ onto $[0, x]$, the function

$$\Phi_k(x, y) := (x, \varphi_k(x, y))$$

maps each rectangle $S_i^k \cap G$ onto $S_i^k \cap \Omega$, and it is bi-Lipschitz on each such rectangle. This allows to define u_k by the implicit equation

$$u_k(x, \varphi_k(x, z)) = \gamma_k(z), \quad (227)$$

and to conclude that $0 \leq u_k \leq 1$ and that u_k is Lipschitz on each $S_i^k \cap \Omega$. Therefore $u_k \in L^\infty \cap BV(\Omega)$, $D_x u_k = D_x^a u_k + D_x^j u_k$ and $D_y u_k = D_y^a u_k$.

Step 2: BV bounds. We prove in this step that $|Du_k|(\Omega)$ is uniformly bounded. This claim and the bound $\|u_k\|_\infty \leq 1$ allow to apply the *BV* compactness theorem to get a subsequence which converges to a bounded *BV* function u , strongly in L^p for every $p < \infty$. In Steps 3 and 4 we will then complete the proof by showing that u satisfies both the requirements (a) and (b).

By differentiating (227) and using (225) we get the following identity for $\mathcal{L}^2$ -a.e. $(x, z) \in S_i^k \cap G$:

$$\begin{aligned} 0 &= \frac{\partial u_k(x, \varphi_k(x, z))}{\partial x} + \frac{\partial u_k(x, \varphi_k(x, z))}{\partial y} \frac{\partial \varphi_k(x, z)}{\partial x} \\ &= \frac{\partial u_k(x, \varphi_k(x, z))}{\partial x} + \frac{\partial u_k(x, \varphi_k(x, z))}{\partial y} \gamma_k(x) \\ &= \frac{\partial u_k(x, \varphi_k(x, z))}{\partial x} + \frac{\partial u_k(x, \varphi_k(x, z))}{\partial y} u_k(x, \varphi_k(x, z)). \end{aligned}$$

Since Φ_k is bi-Lipschitz, we get

$$\partial_x u_k(x, y) + u_k \partial_y u_k(x, y) = 0 \quad \text{for } \mathcal{L}^2\text{-a.e. } (x, y) \in S_i^k \cap \Omega. \quad (228)$$

If $4^{k-1}x \notin \mathbb{N}$ the function $u_k(x, \cdot)$ is nondecreasing. Therefore

$$|D_y u_k|(\Omega) = D_y u_k(\Omega) = \int_1^2 (u_k(x, x) - u_k(x, 0)) dx = 1. \quad (229)$$

From (228) we get

$$|D_x^a u_k|(\Omega) \leq |D_y^a u_k|(\Omega) = 1. \quad (230)$$

Therefore it remains to bound $|D_x^j u_k|(\Omega)$. This consists of

$$|D_x^j u_k|(\Omega) = \sum_{i=1}^{4^{k-1}-1} \int_{V_i^k} |u_k^+ - u_k^-| d\mathcal{H}^1. \quad (231)$$

For each x of type $1 + i4^{1-k}$ we compute

$$\begin{aligned}
 \int_{V_i^k} |u_k^+ - u_k^-| d\mathcal{H}^1 &= \int_0^x |u_k(x^+, y) - u_k(x^-, y)| dy \\
 &= \int_0^1 |\{y: u_k(x^-, y) < t < u_k(x^+, y)\}| dt \\
 &\quad + \int_0^1 |\{y: u_k(x^+, y) < t < u_k(x^-, y)\}| dt \\
 &= \int_0^1 |\{y: u_k(x^-, y) < \gamma_k(z) < u_k(x^+, y)\}| \gamma'_k(z) dz \\
 &\quad + \int_0^1 |\{y: u_k(x^+, y) < \gamma_k(z) < u_k(x^-, y)\}| \gamma'_k(z) dz \\
 &= \int_0^1 |\varphi_k(x^+, z) - \varphi_k(x^-, z)| \gamma'_k(z) dz \\
 &\leq \sup_{z \in]0, 1[} |\varphi_k(x^+, z) - \varphi_k(x^-, z)| \\
 &\stackrel{(224)}{\leq} \frac{4}{3} \sum_{j=1}^k 8^{-j} (\sigma(4^{j-1}x^+) - \sigma(4^{j-1}x^-)). \tag{232}
 \end{aligned}$$

Combining (231) and (232) we get

$$\begin{aligned}
 |D_x^j u_k|(\Omega) &\leq \frac{4}{3} \sum_{i=1}^{4^{k-1}-1} \sum_{j=1}^k 8^{-j} (\sigma(4^{j-1}4^{1-k}i^+) - \sigma(4^{j-1}4^{1-k}i^-)) \\
 &= \frac{4}{3} \sum_{j=1}^k 8^{-j} \sum_{i=1}^{4^{k-1}-1} (\sigma(4^{j-k}i^+) - \sigma(4^{j-k}i^-)) \\
 &= \frac{4}{3} \sum_{j=1}^k 8^{-j} 4^{j-1} \leq \frac{1}{3}. \tag{233}
 \end{aligned}$$

Step 3: Proof of (a). We now fix a bounded BV function u and a subsequence of u_k , not relabeled, which converges to u strongly in L^1 . We claim that (a) holds. More precisely we will show that:

(CI) For $\mathcal{L}^1$ -a.e. x the function $u(x, \cdot)$ is a nonconstant BV function of one variable which has no absolutely continuous part and no jump part.

(CI) gives (a) by the slicing theory of BV functions, see Theorem 3.108 of [11].

In order to prove (CI) we proceed as follows. By possibly extracting another subsequence we assume that u_k converges to u $\mathcal{L}^2$ -a.e. in Ω . We then show (CI) for every x such that:

- $4^k x \notin \mathbb{N}$ for every k ;
- $u_k(x, y)$ converges to $u(x, y)$ for $\mathcal{L}^1$ -a.e. y .

Clearly $\mathcal{L}^1$ -a.e. x meets these requirements.

Fix any such x . Note that x is never on the boundary of any strip S_i^k . Therefore we can denote by g_k^x the inverse of $\varphi_k(x, \cdot)$ and we can use (227) to write

$$u_k(x, y) = \gamma_k(g_k^x(y)). \quad (234)$$

Thanks to (226), the Lipschitz constant of g_k is uniformly bounded. Therefore, after possibly extracting a subsequence, we can assume that g_k uniformly converge to a Lipschitz function g . Since γ_k uniformly converge to the Cantor ternary function γ , we can pass into the limit in (234) to conclude

$$u(x, y) = \gamma(g(y)). \quad (235)$$

Therefore $u(x, \cdot)$ is continuous, nondecreasing, nonconstant, and locally constant outside a closed set of zero Lebesgue measure ($g^{-1}(C)$, where C is the Cantor set). This proves (CI).

Step 4: Proof of (b). Let u be as in Step 3. From the construction of u_k it follows that

$$D_x u_k + D_y \left(\frac{u_k^2}{2} \right) = D_x^j u_k. \quad (236)$$

After possibly extracting a subsequence we can assume that $D_x^j u_k$ converges weakly* to a measure μ . This gives

$$D_x u + D_y \left(\frac{u^2}{2} \right) = \mu. \quad (237)$$

Therefore it suffices to prove that μ is concentrated on a set of σ -finite one-dimensional Hausdorff measure. Indeed μ is concentrated on the union of the countable family of segments $\{V_{k,i}^l\}$. In order to prove this claim it suffices to show the following tightness property: for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|D_x^j u_k| \left(\bigcup_{l \geq N} \bigcup_{i=1}^{4^{l-1}-1} V_i^l \right) \leq \varepsilon \quad \text{for every } k. \quad (238)$$

Note that

$$|D_x^j u_k| \left(\bigcup_{l \geq N} \bigcup_{i=1}^{4^{l-1}-1} V_i^l \right) \leq \sum_{l \geq N} \sum_{i=1}^{4^{l-1}-1} \int_{V_i^l} |u_k^+ - u_k^-|.$$

Then the same computations leading to (232) and (233) give

$$|D_x^j u_k| \left(\bigcup_{l \geq N} \bigcup_{j=1}^{4^{l-1}-1} v_j^l \right) \leq \frac{4}{3} \sum_{l=N}^k 8^{-l} 4^{l-1} \leq \frac{1}{3 \cdot 2^{N-1}}. \quad (239)$$

This concludes the proof. $\square$

REMARK 9.2. The function u constructed in Proposition 9.1 solves Burgers' equation with a measure source

$$D_t u + D_x \left(\frac{u^2}{2} \right) = \mu, \quad (240)$$

and has nonvanishing Cantor part. On the other hand, in [9] it has been proved that *entropy* solutions to Burgers' equation without source are *SBV*, i.e., the Cantor part of their derivative is trivial. It would be interesting to understand whether this gain of regularity is due to the entropy condition, or instead *BV* distributional solutions of (240) with $\mu = 0$ are always *SBV*.

Acknowledgements

This research has been partially supported by the Swiss National Foundation. The first part of these notes are based on the Cours Poupaud that the author held in March 2006 in the University of Nice. He wishes to thank the organizers and the audience, in particular Alessio Figalli, for pointing out many mistakes in the first drafts.

References

- [1] G. Alberti, *Rank-one properties for derivatives of functions with bounded variations*, Proc. Roy. Soc. Edinburgh Sect. A **123** (1993), 239–274.
- [2] L. Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math. **158** (2004), 227–260.
- [3] L. Ambrosio, *Transport equation and Cauchy problem for non-smooth vector fields*, Lect. Notes CIME Summer School in Cetraro, June 27–July 2, 2005, Preprint (2005), available at <http://cvgmt.sns.it/cgi/get.cgi/papers/amb05/>.
- [4] L. Ambrosio, F. Bouchut and C. De Lellis, *Well-posedness for a class of hyperbolic systems of conservation laws in several space dimensions*, Comm. Partial Differential Equations **29** (2004), 1635–1651.
- [5] L. Ambrosio and G. Crippa, *Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields*, Proc. School “Multi-D Hyperbolic Conservation Laws” (Bologna, January 17–20, 2005), to appear. Preprint, available at <http://cvgmt.sns.it/cgi/get.cgi/papers/ambcri06/>.
- [6] L. Ambrosio, G. Crippa and S. Maniglia, *Traces and fine properties of a BD class of vector fields and applications*, Ann. Fac. Sci. Toulouse Math. (6) **14** (2005), 527–561. Available at <http://cvgmt.sns.it/cgi/get.cgi/papers/ambcriman04/>.
- [7] L. Ambrosio and G. Dal Maso, *A general chain rule for distributional derivatives*, Proc. Amer. Math. Soc. **108** (1990), 691–702.

- [8] L. Ambrosio and C. De Lellis, *Existence of solutions for a class of hyperbolic systems of conservation laws in several space dimensions*, Int. Math. Res. Not. **41** (2003), 2205D2220.
- [9] L. Ambrosio and C. De Lellis, *A note on admissible solutions of 1d scalar conservation laws and 2d Hamilton–Jacobi equations*, J. Hyperbolic Differential Equations **1** (4) (2004), 813D826.
- [10] L. Ambrosio, C. De Lellis and J. Mal., *On the chain rule for the divergence of vector fields: Applications, partial results, open problems*, Perspectives in Nonlinear Partial Differential Equations: In Honor of Haim Brezis, to appear. Preprint, available at <http://cvgmt.sns.it/papers/ambdel05/>.
- [11] L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Math. Monogr. Clarendon Press/Oxford Univ. Press, New York (2000).
- [12] G. Anzellotti, *Pairings between measures and bounded functions and compensated compactness*, Ann. Math. Pura Appl. **135** (1983), 293D318.
- [13] G. Anzellotti, *Traces of bounded vector fields and the divergence theorem*, Unpublished preprint (1983).
- [14] G. Anzellotti, *The Euler equation for functionals with linear growth*, Trans. Amer. Math. Soc. **290** (1985), 483D501.
- [15] F. Bouchut, *Renormalized solutions to the Vlasov equation with coefficients of bounded variation*, Arch. Ration. Mech. Anal. **157** (2001), 75D90.
- [16] P. Brenner, *The Cauchy problem for the symmetric hyperbolic systems in L^p* , Math. Scand. **19** (1966), 27D37.
- [17] A. Bressan, *An ill posed Cauchy problem for a hyperbolic system in two space dimensions*, Rend. Sem. Mat. Univ. Padova **110** (2003), 103D117.
- [18] A. Bressan, *A lemma and a conjecture on the cost of rearrangements*, Rend. Sem. Mat. Univ. Padova **110** (2003), 97D102.
- [19] A. Bressan, *Some remarks on multidimensional systems of conservation laws*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. **15** (2004), 225D233.
- [20] G.Q. Chen and H. Frid, *Divergence-measure fields and conservation laws*, Arch. Ration. Mech. Anal. **147** (1999), 89D118.
- [21] G.Q. Chen and H. Frid, *Extended divergence-measure fields and the Euler equation of gas dynamics*, Comm. Math. Phys. **236** (2003), 251D280.
- [22] C. Dafermos, *Stability for systems of conservation laws in several space dimensions*, SIAM J. Math. Anal. **26** (1995), 1403D1414.
- [23] C. Dafermos, *Hyperbolic Conservation Laws in Continuum Physics*, Grundlehren Math. Wiss., Vol. 325, Springer-Verlag, Berlin (2000).
- [24] E. De Giorgi and L. Ambrosio, *Un nuovo funzionale del calcolo delle variazioni*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (8) Mat. Appl. **82** (1988), 199D210.
- [25] C. De Lellis, *Blow-up of the BV norm in the multidimensional Keyfitz and Kranzer system*, Duke Math. J. **127** (2005), 313D339.
- [26] C. De Lellis, *A note on Alberti’s rank-one theorem*, Proc. School “Multi-D Hyperbolic Conservation Laws” (Bologna, January 17D20, 2005), to appear. Available at <http://cvgmt.sns.it/cgi/get.cgi/papers/delb/>.
- [27] R. DiPerna and P.L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math. **98** (1989), 511D517.
- [28] H. Freistühler, *A standard model of generic rotational degeneracy*, Nonlinear Hyperbolic Equations D Theory, Computation Methods, and Applications, Aachen, 1988, Notes Numer. Fluid Mech., Vol. 24, Vieweg, Braunschweig (1989), 149D158.
- [29] H. Freistühler, *Rotational degeneracy of hyperbolic systems of conservation laws*, Arch. Ration. Mech. Anal. **113** (1990), 39D64.
- [30] H. Freistühler, *Non-uniformity of vanishing viscosity approximation*, Appl. Math. Lett. **6** (2) (1993), 35D41.
- [31] H. Freistühler, *Dynamical stability and vanishing viscosity: A case study of a non-strictly hyperbolic system*, Comm. Pure Appl. Math. **45** (5) (1992), 561D582.
- [32] H. Frid, *Asymptotic stability of non-planar Riemann solutions for a special class of multi-d systems of conservation laws*, J. Hyperbolic Differential Equation **1** (2004), 567D579.
- [33] E. Gagliardo, *Caratterizzazione delle tracce sull la frontiera relative ad alcune classi di funzioni in più variabili*, Rend. Sem. Mat. Univ. Padova **27** (1957), 284D305.
- [34] B.L. Keyfitz and H.C. Kranzer, *A system of nonstrictly hyperbolic conservation laws arising in elasticity theory*, Arch. Ration. Mech. Anal. **72** (1980), 219D241.

- [35] B.L. Keyfitz and C.A. Mora, *Prototypes for nonstrict hyperbolicity in conservation laws*, Nonlinear PDEs, Dynamics and Continuum Physics (South Hadley, MA, 1998), Contemp. Math., Vol. 255, Amer. Math. Soc., Providence, RI (2000), 125–137.
- [36] S. Kruzhkov, *First-order quasilinear equations with several space variables*, Math. USSR Sb. **10** (1970), 217–273.
- [37] J. Rauch, *BV estimates fail for most quasilinear systems in dimension greater than one*, Comm. Math. Phys. **106** (1986), 481–484.
- [38] D. Serre, *System of conservation laws I, II*, Cambridge Univ. Press, Cambridge (1999).
- [39] A. Vasseur, *Strong traces for solutions of multidimensional scalar conservation laws*, Arch. Ration. Mech. Anal. **160** (2001), 181–193.

CHAPTER 5

Collisionless Kinetic Equations from Astrophysics – The Vlasov-Poisson System

Gerhard Rein

Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany

E-mail: gerhard.rein@uni-bayreuth.de

Contents

Introduction	385
Notation and preliminaries	387
1. Classical solutions to the initial value problem	390
1.1. The initial value problem – an overview	390
1.2. Local existence	393
1.3. Spherically symmetric solutions	402
1.4. Small data solutions	408
1.5. Conservation laws and a priori bounds	413
1.6. Global existence for general data	417
1.7. Asymptotic behavior	431
2. Stability	435
2.1. Introduction – steady states, stability and energy-Casimir functionals	435
2.2. Reduction	443
2.3. Existence of minimizers via the reduced problem	449
2.4. Existence of minimizers – the direct approach	455
2.5. Minimizers are steady states	460
2.6. Dynamical stability	465
2.7. The reduced variational problem and the Euler-Poisson system	470
References	472

Introduction

Many important developments and concepts in mathematics originate with the N -body problem. It describes the motion of N mass points which move according to Newton's equations of motion under the influence of their mutual attraction governed by Newton's law of gravity. The N -body problem has many applications in astronomy and astrophysics, the most notable one being our solar system. Looking at larger astronomical scales further N -body systems come into view, for example globular clusters or galaxies. If the internal structure of the stars, interstellar material, processes leading to the birth or death of stars, and various other effects are neglected, then a galaxy can be described as an N -body system. For the solar system N is a fairly small number, and it is to an excellent degree of precision possible to predict the exact positions of all N bodies. For a galaxy however, N is of the order of 10^{10} – 10^{12} , and keeping track of all these mass points is neither feasible nor even desired. Instead, the evolution of the in some sense averaged mass distribution of the galaxy is the issue. Such a statistical description of a large ensemble of gravitationally interacting mass points leads to a mathematical problem which is far more tractable in certain of its aspects than the N -body problem is even for very moderate N . In the present treatise results for certain nonlinear systems of partial differential equations are presented, which are used in the modeling of galaxies, globular clusters, and many other systems where a large ensemble of mass points interacts by a force field which the ensemble creates collectively.

In order to motivate the equations which describe such a particle ensemble let us continue to think of a galaxy. If $U = U(t, x)$ denotes its gravitational potential depending on time $t \in \mathbb{R}$ and position $x \in \mathbb{R}^3$, then an individual star of unit mass with position x and velocity $v \in \mathbb{R}^3$ obeys Newton's equations of motion

$$\dot{x} = v, \quad \dot{v} = -\partial_x U(t, x), \quad (1)$$

as long as it has no close encounters with other stars. Here $\partial_x U$ denotes the gradient of U with respect to x . To describe the galaxy as a whole we introduce its density $f = f(t, x, v) \geq 0$ on phase space $\mathbb{R}^3 \times \mathbb{R}^3$. The integral of f over any region of phase space gives the mass or number of particles (stars) which at that instant of time have phase space coordinates in that region. In a typical galaxy collisions among stars are sufficiently rare to be (in a first approximation) negligible. Hence f is constant along solutions of the equations of motion (1) and satisfies a first-order conservation law on phase space, the characteristic system of which are the equations of motion (1) of a single test particle,

$$\partial_t f + v \cdot \partial_x f - \partial_x U \cdot \partial_v f = 0. \quad (2)$$

Of course this equation can be derived in a more rigorous way like other conservation laws, using Gauss' theorem. The spatial mass density $\rho = \rho(t, x)$ induced by f determines the gravitational potential U according to Newton's law for gravity, subject to the

usual boundary condition at spatial infinity,

$$\Delta U = 4\pi\gamma\rho, \quad \lim_{|x| \rightarrow \infty} U(t, x) = 0, \quad (3)$$

$$\rho(t, x) = \int f(t, x, v) dv; \quad (4)$$

for the moment $\gamma = 1$. Equations (2)–(4) form a closed, nonlinear system of partial differential equations which governs the time evolution of a self-gravitating collisionless ensemble of particles. No additional equations such as an equation of state are needed to close this system, as would be the case for fluid type models. A more detailed discussion of the derivation of such a system and its underlying physical assumptions can be found in [73].

At the beginning of the last century the astrophysicist Sir J. Jeans used this system to model stellar clusters and galaxies [65] and to study their stability properties. In this context it appears in many textbooks on astrophysics such as [13,26]. If we want to model an ensemble of mass points which interact by a repulsive electrostatic potential, we choose $\gamma = -1$. This form of the system is important in plasma physics where it was introduced by A.A. Vlasov around 1937 [109,110]. In the mathematics literature the system of equations (2)–(4) has become known as the Vlasov–Poisson system.

Besides being nonlinear the specific mathematical difficulty of this system lies in the fact that an equation on phase space is coupled to an equation on space. The Vlasov equation easily provides a priori bounds on L^p -norms of $f(t)$ for any $p \in [1, \infty]$, but upon integration with respect to v only an L^1 -bound on $\rho(t)$ survives, which does not give good bounds for $\partial_x U$.

The Vlasov–Poisson system is just one example of a class of partial differential equations known as kinetic equations. Other such systems are the Vlasov–Maxwell system, a Lorentz invariant model for a dilute plasma where the particles interact by electrodynamic fields, and the Vlasov–Einstein system which describes a self-gravitating collisionless ensemble of mass points in the framework of General Relativity; for more details we refer to Section 1.1. In these systems the standing assumptions are that the particle ensemble is sufficiently large to justify a description by a (smooth) density function on phase space and that collisions are sufficiently rare to be neglected. If collisions are to be included a Boltzmann collision operator replaces the zero on the right-hand side of the Vlasov equation (2). One can then consider situations where collisions are the only interaction among the particles, the case of the classical Boltzmann equation of gas dynamics, or situations where both short and long range interactions are taken into account, like in the Vlasov–Poisson–Boltzmann or Vlasov–Maxwell–Boltzmann systems.

We refer to [18,22,28,43,44] for systems including collisions. The present treatise is concerned with the collisionless case. We essentially consider two topics: The existence of classical, smooth solutions to the initial value problem, and the nonlinear stability of stationary solutions. In dealing with these problems we focus on the Vlasov–Poisson system, for the stability problem we restrict ourselves even further and consider only the gravitational case. The motivation for this approach is as follows. For the Vlasov–Poisson system the mathematical understanding of the initial value problem is fairly complete, while on

the other hand the techniques which were successful there can provide a guide to attacking open problems for related systems. For the stability problem in the gravitational case, which as noted above was one of the starting points of the whole field, a successful approach began to appear in the last few years, with techniques which hopefully will reach beyond kinetic theory. In spite of the restriction to the Vlasov–Poisson system we frequently comment on results and open problems for related systems so that this treatise can serve as a guide into the whole field of collisionless kinetic equations.

There are important questions concerning kinetic equations and even concerning the Vlasov–Poisson system which we do not discuss. An obvious one, which comes to mind in connection with our point of departure, the N -body problem, is the following: Consider a sequence of N -body problems where N increases to infinity and where the initial data, which can be interpreted as sums of Dirac δ distributions on phase space, converge in an appropriate sense to a smooth initial distribution function on phase space. Do the solutions of the N -body problems at later times then converge to the solution of the Vlasov–Poisson system launched by this initial distribution? A positive answer to this question could be considered as a rigorous derivation of the Vlasov–Poisson system from the N -body problem, but the question is open. Partial results, where the Newtonian interaction potential $1/|x|$ is replaced by less singular ones, are given in [54,85].

Notation and preliminaries

Our notation is mostly standard or self-explaining, but to avoid misunderstandings we fix some of it here. For $x, y \in \mathbb{R}^n$ the Euclidean scalar product and norm are denoted by

$$x \cdot y := \sum_{i=1}^n x_i y_i, \quad |x| := \sqrt{x \cdot x}.$$

The open ball of radius $R > 0$ with center $x \in \mathbb{R}^n$ is denoted by

$$B_R(x) := \{y \in \mathbb{R}^n \mid |x - y| < R\}, \quad B_R := B_R(0).$$

For $\xi \in \mathbb{R}$ the positive part of this number is

$$\xi_+ := \max\{\xi, 0\}.$$

For a set $M \subset \mathbb{R}^n$, $\mathbb{1}_M$ denotes its indicator function,

$$\mathbb{1}_M(x) = 1 \text{ if } x \in M, \quad \mathbb{1}_M(x) = 0 \text{ if } x \notin M.$$

For a differentiable function $f = f(t, x, v)$, $t \in \mathbb{R}$, $x, v \in \mathbb{R}^3$,

$$\partial_t f, \quad \partial_x f, \quad \partial_v f$$

denote its partial derivatives with respect to the indicated variable; in the case of x or v these are actually gradients. If $U = U(x)$ we also write $\nabla U = \partial_x U$ for the gradient. For

$t \in \mathbb{R}$ we denote by $f(t)$ the function

$$f(t) : \mathbb{R}^3 \times \mathbb{R}^3 \ni (x, v) \mapsto f(t, x, v).$$

By

$$C^k(\mathbb{R}^n), \quad C_c^k(\mathbb{R}^n)$$

we denote the space of k times continuously differentiable functions on $\mathbb{R}^n$, the subscript “c” indicates compactly supported functions. The Lebesgue measure of a measurable set $M \subset \mathbb{R}^n$ is denoted by $\text{vol}(M)$. The norm on the usual Lebesgue spaces $L^p(\mathbb{R}^n)$, $1 \leq p \leq \infty$, is denoted by $\|\cdot\|_p$, where by default the corresponding integral extends over $\mathbb{R}^n$ with $n = 3$ or $n = 6$ as the case may be. We denote by $L_+^1(\mathbb{R}^n)$ the set of nonnegative integrable functions on $\mathbb{R}^n$. For $f \in L_+^1(\mathbb{R}^6)$ we define the induced spatial density $\rho_f \in L_+^1(\mathbb{R}^3)$ by

$$\rho_f(x) := \int f(x, v) dv;$$

integrals without explicitly specified domain of integration always extend over $\mathbb{R}^3$. For $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$ measurable the induced potential is denoted by

$$U_\rho(x) := -\gamma \int \frac{\rho(y)}{|x - y|} dy,$$

provided the latter convolution integral exists, also $U_f := U_{\rho_f}$, and if $f = f(t, x, v)$ or $\rho = \rho(t, x)$ also depend on time t we write $\rho_f(t, x) := \rho_{f(t)}(x)$, $U_\rho(t, x) := U_{\rho(t)}(x)$ etc. It will often not be necessary to write the subscripts at all.

For the convenience of the reader we collect some facts from potential theory.

LEMMA P1. *Let $\rho \in C_c^1(\mathbb{R}^3)$. Then the following holds:*

(a) U_ρ is the unique solution of

$$\Delta U = 4\pi\gamma\rho, \quad \lim_{|x| \rightarrow \infty} U(x) = 0$$

in $C^2(\mathbb{R}^3)$. Moreover,

$$\nabla U_\rho(x) = \gamma \int \frac{x - y}{|x - y|^3} \rho(y) dy,$$

$$U_\rho(x) = O(|x|^{-1}), \quad \nabla U_\rho(x) = O(|x|^{-2}) \quad \text{for } |x| \rightarrow \infty.$$

(b) For any $p \in [1, 3[$,

$$\|\nabla U_\rho\|_\infty \leq c_p \|\rho\|_p^{p/3} \|\rho\|_\infty^{1-p/3},$$

where the constant $c_p > 0$ depends only on p , in particular, $c_1 = 3(2\pi)^{2/3}$. Moreover, the second-order derivative satisfies, for any $0 < d \leq R$,

$$\|D^2 U_\rho\|_\infty \leq c \left[R^{-3} \|\rho\|_1 + d \|\nabla \rho\|_\infty + \left(1 + \ln \frac{R}{d}\right) \|\rho\|_\infty \right],$$

with $c > 0$ independent of ρ , R , d , and

$$\|D^2 U_\rho\|_\infty \leq c \left[(1 + \|\rho\|_\infty)(1 + \ln_+ \|\nabla \rho\|_\infty) + \|\rho\|_1 \right].$$

PROOF. We only sketch the proof since most of this is well known. The formula for ∇U_ρ is obtained by shifting the x -variable into the argument of ρ first. After differentiating under the integral once the derivative can be moved from ρ to the kernel $1/|x - y|$ using Gauss' theorem; one has to exclude a small ball of radius ε about the singularity $y = x$ when doing so, but the corresponding boundary term vanishes as $\varepsilon \rightarrow 0$. This procedure is then applied to the formula for ∇U_ρ , except that now the boundary term survives in the limit $\varepsilon \rightarrow 0$, and the resulting singularity is no longer integrable at $y = x$. Hence for any $d > 0$ and $i, j = 1, 2, 3$,

$$\begin{aligned} \partial_{x_i} \partial_{x_j} U_\rho(x) &= -\gamma \int_{|x-y| \geq d} \left[3 \frac{(x_i - y_i)(x_j - y_j)}{|x - y|^5} - \frac{\delta_{ij}}{|x - y|^3} \right] \rho(y) dy \\ &\quad - \gamma \int_{|x-y| \leq d} [\cdots] (\rho(y) - \rho(x)) dy + \frac{4\pi}{3} \gamma \delta_{ij} \rho(x); \end{aligned}$$

the difference $\rho(y) - \rho(x)$ in the latter integral kills one power of the singularity so the integral exists. The uniqueness assertion is usually referred to as Liouville's theorem, and the asymptotic behavior is easy to deduce from the compact support of ρ . We consider (b) in more detail. For any $R > 0$, Hölder's inequality implies that

$$\begin{aligned} |\nabla U_\rho(x)| &\leq \int_{|x-y| < R} \frac{|\rho(y)|}{|x - y|^2} dy + \int_{|x-y| \geq R} \frac{|\rho(y)|}{|x - y|^2} dy \\ &\leq 4\pi R \|\rho\|_\infty + \left(\frac{4\pi}{2q - 3} R^{3-2q} \right)^{1/q} \|\rho\|_p, \end{aligned}$$

where $1/p + 1/q = 1$ and hence $q > 3/2$. We optimize this estimate by choosing $R = (c \|\rho\|_p / \|\rho\|_\infty)^{p/3}$ with a suitable constant $c > 0$ and obtain the estimate for ∇U_ρ . With $0 < d \leq R$ the formula for $\partial_{x_i} \partial_{x_j} U_\rho$ implies that

$$\begin{aligned} |D^2 U_\rho(x)| &\leq \frac{4\pi}{3} |\rho(x)| + \|\nabla \rho\|_\infty \int_{|x-y| \leq d} \frac{4}{|x - y|^2} dy \\ &\quad + \int_{d < |x-y| \leq R} \frac{4}{|x - y|^3} |\rho(y)| dy + \int_{|x-y| > R} \frac{4}{|x - y|^3} |\rho(y)| dy \\ &\leq c \left[\|\rho\|_\infty + d \|\nabla \rho\|_\infty + \|\rho\|_\infty \ln \frac{R}{d} + R^{-3} \|\rho\|_1 \right]. \end{aligned}$$

The second form of the estimate results by the choice $R = 1$, and $d = 1/\|\nabla\rho\|_\infty$ if $\|\nabla\rho\|_\infty \geq 1$, else $d = 1$. $\square$

We also need certain L^p -estimates for the potential, based on the weak Young's inequality.

LEMMA P2. (a) Let $1 < p, q, r < \infty$ with $1/p + 1/q = 1 + 1/r$. Then for all functions $g \in L^p(\mathbb{R}^n)$, $h \in L_w^q(\mathbb{R}^n)$ the convolution $g * h := \int g(\cdot - y)h(y) dy \in L^r(\mathbb{R}^n)$ satisfies

$$\|g * h\|_r \leq c \|g\|_p \|h\|_{q,w}.$$

Here $c = c(p, q, n) > 0$, and by definition, $h \in L_w^q(\mathbb{R}^n)$ iff h is measurable and

$$\|h\|_{q,w} := \sup_{\tau > 0} \tau (\text{vol}\{x \in \mathbb{R}^n \mid |h(x)| > \tau\})^{1/q} < \infty;$$

the latter expression does not define a norm.

(b) If $\rho \in L^{6/5}(\mathbb{R}^3)$ then $U_\rho \in L^6(\mathbb{R}^3)$ with weak derivative $\nabla U_\rho := \gamma \cdot / |\cdot|^3 * \rho \in L^2(\mathbb{R}^3)$, and there exists a constant $c > 0$ such that

$$\frac{1}{8\pi} \int |\nabla U_\rho|^2 dx = \frac{1}{2} \iint \frac{\rho(x)\rho(y)}{|x-y|} dx dy \leq c \|\rho\|_{6/5}^2.$$

PROOF. As to the weak Young's inequality recalled in (a), cf. [74], Section 4.3. The assertions for U_ρ and ∇U_ρ follow with $n = 3$, $p = 6/5$, and $h = |\cdot|^{-1}$, $q = 3$ or $h = |\cdot|^{-2}$, $q = 3/2$. The estimate in (b), a special case of the Hardy–Littlewood–Sobolev inequality, follows by Hölder's inequality. If $\rho \in C_c^1(\mathbb{R}^3)$ then $\gamma \cdot / |\cdot|^3 * \rho$ is the gradient of U_ρ , and integration by parts together with the Poisson equation yields the equality of the two integrals. The general case follows by a density argument. $\square$

1. Classical solutions to the initial value problem

1.1. The initial value problem – an overview

Before going into details we give an overview of this chapter and of the history and current state of the mathematical treatment of the initial value problem for collisionless kinetic systems. We also introduce some systems which are related to the Vlasov–Poisson system

$$\partial_t f + v \cdot \partial_x f - \partial_x U \cdot \partial_v f = 0, \quad (1.1)$$

$$\Delta U = 4\pi\gamma\rho, \quad \lim_{|x| \rightarrow \infty} U(t, x) = 0, \quad (1.2)$$

$$\rho(t, x) = \int f(t, x, v) dv; \quad (1.3)$$

as a rule, $t \in \mathbb{R}$, $x, v \in \mathbb{R}^3$. If the particles are allowed to move at relativistic speeds a first modification is to replace the Vlasov equation (1.1) by

$$\partial_t f + \frac{v}{\sqrt{1 + |v|^2}} \cdot \partial_x f - \partial_x U \cdot \partial_v f = 0. \quad (1.4)$$

Here v should be viewed as momentum so that $v/\sqrt{1 + |v|^2}$ is the corresponding relativistic velocity; like all other physical constants the speed of light is normalized to unity. The system (1.4), (1.2), (1.3) is called the relativistic Vlasov–Poisson system, and again one distinguishes the gravitational case $\gamma = 1$ and the plasma physics case $\gamma = -1$. In spite of its name this system is not fully relativistic, i.e., not Lorentz invariant. To obtain a Lorentz invariant system the field equation has to be modified accordingly. In the plasma physics case this yields the relativistic Vlasov–Maxwell system which we write for a plasma with two particle species of opposite charge; otherwise, all physical constants are again normalized to unity:

$$\begin{aligned} \partial_t f^\pm + \frac{v}{\sqrt{1 + |v|^2}} \cdot \partial_x f^\pm \pm \left(E + \frac{v}{\sqrt{1 + |v|^2}} \times B \right) \cdot \partial_v f^\pm &= 0, \\ \partial_t E - \operatorname{curl} B &= -4\pi j, \quad \partial_t B + \operatorname{curl} E = 0, \\ \operatorname{div} E &= 4\pi \rho, \quad \operatorname{div} B = 0, \\ \rho(t, x) &= \int (f^+ - f^-)(t, x, v) dv, \\ j(t, x) &= \int \frac{v}{\sqrt{1 + |v|^2}} (f^+ - f^-)(t, x, v) dv. \end{aligned}$$

Here $f^\pm = f^\pm(t, x, v)$ are the densities of the positively or negatively charged particles on phase space respectively, $E = E(t, x)$ and $B = B(t, x)$ denote the electric and magnetic field, and the source terms in the Maxwell field equations are the charge and current density $\rho = \rho(t, x)$ and $j = j(t, x)$. According to [71], p. 124, it was this system (with v instead of $v/\sqrt{1 + |v|^2}$) which Vlasov introduced into the plasma physics literature in 1937. The major assumption that collisions can be neglected is satisfied if the plasma is very hot and/or very dilute, a good example being the solar wind. If a fully relativistic description in the gravitational case is desired the corresponding Vlasov equation has to be coupled to Einstein's field equations

$$G^{\alpha\beta} = 8\pi T^{\alpha\beta}.$$

Here the Einstein tensor $G^{\alpha\beta}$ is a nonlinear second-order differential expression in terms of the Lorentz metric $g_{\alpha\beta}$ on the space–time manifold M , and $T^{\alpha\beta}$ is the energy momentum tensor. The equations of motion of a test particle are the geodesic equations in the metric $g_{\alpha\beta}$ so that the corresponding Vlasov equation is that first-order differential equation on the tangent bundle TM of the space–time manifold M which has the geodesic equations as its characteristic system. The corresponding density f on phase space TM determines

the energy momentum tensor $T^{\alpha\beta}$. For the present treatise there is no need to make this more precise, and we refer to [3] for an introduction to the Vlasov–Einstein system.

We first consider classical solutions of the Vlasov–Poisson system (1.1)–(1.3), i.e., solutions where all relevant derivatives exist in the classical sense. A local existence and uniqueness result to the initial value problem was established by Kurth [69]. A first global existence result was proven by Batt [6] for a modified system where the spatial density ρ is regularized. The first global existence result for the original problem was again obtained by Batt [7] for spherically symmetric data. In the course of this proof an important continuation criterion was established: A local solution can be extended as long as its velocity support is under control. In Section 1.2 we prove the local existence result together with this continuation criterion, since it forms the basis for results toward global existence. The analogous result is valid for the relativistic Vlasov–Poisson and Vlasov–Maxwell systems [36], see also [15,66], and for the Vlasov–Einstein system in the spherically symmetric, asymptotically flat case [100]. In Section 1.3 spherical symmetry is shown to imply global existence for the Vlasov–Poisson system. It is also shown to be essential that the particle distribution is given by a regular function on phase space; if the particles are allowed to be δ -distributed in velocity space blow-up in finite time can occur.

The next major step was a global existence result for the Vlasov–Poisson system with sufficiently small data by Bardos and Degond [5]. The analogous result was achieved for the Vlasov–Maxwell system by Glassey and Strauss [37], and for the Vlasov–Einstein system in the spherically symmetric, asymptotically flat case by Rein and Rendall [100]. The corresponding techniques are discussed in Section 1.4 for the case of the Vlasov–Poisson system.

The development for the Vlasov–Poisson system culminated in 1989 when independently and almost simultaneously two different proofs for global existence of classical solutions for general data were given, one by Pfaffelmoser [89] and one by Lions and Perthame [80]. Since the two approaches are quite different from each other and both have their strengths we present them both in Section 1.6. In sharp contrast to the N -body problem global existence is obtained both for the repulsive and for the attractive case. In the former case the total energy is positive definite while in the latter it is indefinite, but as shown in Section 1.5, the same a priori bounds can be derived in both cases.

It may seem strange to discuss results for spherically symmetric and for small initial data when there is a result for general ones. The reason is that for the restricted data more information on the behavior of the solution is obtained and, more importantly, the techniques employed may be useful for similar problems where a general result is not yet available.

For the Vlasov–Maxwell system no analogous global existence result for general data has been proven yet. However, two points have to be emphasized here: Firstly, our discussion so far refers to the full three-dimensional problem, and much progress has been made for lower-dimensional versions of the Vlasov–Maxwell system [31–33]. Secondly, our discussion so far is restricted to classical, smooth solutions. In a celebrated paper R. DiPerna and P.-L. Lions proved global existence of appropriately defined weak solutions for the Vlasov–Maxwell system, cf. [21]. A somewhat simplified proof of this result under somewhat more restrictive assumptions can be found in [99]. A variety of tools for kinetic equations, which are used in these results and in many others, is discussed in a much broader context in [88].

Granted that global existence holds for both the attractive and the repulsive case of the Vlasov–Poisson system one certainly expects a different behavior of the two cases for large times. Results in this direction were obtained in [24,63,87], and they are discussed in Section 1.7.

1.2. Local existence

A local existence and uniqueness theorem is the necessary starting point for all further investigations. Since the basic approach to proving such a result is used for many related systems again and again, it is worthwhile to give a complete proof here. To begin with we make precise what we mean by a classical solution:

DEFINITION. A function $f : I \times \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow [0, \infty[$ is a classical solution of the Vlasov–Poisson system on the interval $I \subset \mathbb{R}$ if the following holds:

- (i) The function f is continuously differentiable with respect to all its variables.
- (ii) The induced spatial density $\rho = \rho_f$ and potential $U = U_f$ exist on $I \times \mathbb{R}^3$. They are continuously differentiable, and U is twice continuously differentiable with respect to x .
- (iii) For every compact subinterval $J \subset I$ the field $\partial_x U$ is bounded on $J \times \mathbb{R}^3$.
- (iv) The functions f, ρ, U satisfy the Vlasov–Poisson system (1.1)–(1.3) on $I \times \mathbb{R}^3 \times \mathbb{R}^3$.

It is essential that the local existence result not only provides unique local solutions for a sufficiently large class of initial data, but also says in which way a solution can possibly stop to exist after a finite time.

THEOREM 1.1. Every initial datum $\mathring{f} \in C_c^1(\mathbb{R}^6)$, $\mathring{f} \geq 0$, launches a unique classical solution f on some time interval $[0, T[$ with $f(0) = \mathring{f}$. For all $t \in [0, T[$ the function $f(t)$ is compactly supported and nonnegative. If $T > 0$ is chosen maximal and if

$$\sup\{|v| \mid (x, v) \in \text{supp } f(t), 0 \leq t < T\} < \infty$$

or

$$\sup\{\rho(t, x) \mid 0 \leq t < T, x \in \mathbb{R}^3\} < \infty,$$

then the solution is global, i.e., $T = \infty$.

A classical solution can be extended as long as its velocity support or its spatial density remain bounded. This rules out a breakdown of the solution by shock formation where typically the solution remains bounded but a derivative blows up; if the solution blows up, ρ must blow up due to a concentration effect.

Due to the requirement that the initial datum be compactly supported, the solutions obtained in the theorem enjoy stronger properties than what is required in the definition. This requirement can be replaced by suitable fall-off conditions at infinity [57]. Since such an extension of the result is mostly technical we adopt the simpler case. One can weaken the

requirements on a solution and still retain uniqueness. Exactly how far one can weaken the solution concept without losing uniqueness is not yet completely understood.

In order to prove the theorem we need to be able to solve the Vlasov equation for a given field $F = -\partial_x U$. First we consider the characteristic flow.

LEMMA 1.2. *Let $I \subset \mathbb{R}$ be an interval and let $F \in C(I \times \mathbb{R}^3; \mathbb{R}^3)$ be continuously differentiable with respect to x and bounded on $J \times \mathbb{R}^3$ for every compact subinterval $J \subset I$. Then for every $t \in I$ and $z = (x, v) \in \mathbb{R}^3 \times \mathbb{R}^3$ there exists a unique solution $I \ni s \mapsto (X, V)(s, t, x, v)$ of the characteristic system*

$$\dot{x} = v, \quad \dot{v} = F(s, x) \quad (1.5)$$

with $(X, V)(t, t, x, v) = (x, v)$. The characteristic flow $Z := (X, V)$ has the following properties:

- (a) $Z : I \times I \times \mathbb{R}^6 \rightarrow \mathbb{R}^6$ is continuously differentiable.
- (b) For all $s, t \in I$ the mapping $Z(s, t, \cdot) : \mathbb{R}^6 \rightarrow \mathbb{R}^6$ is a C^1 -diffeomorphism with inverse $Z(t, s, \cdot)$, and $Z(s, t, \cdot)$ is measure preserving, i.e.,

$$\det \frac{\partial Z}{\partial z}(s, t, z) = 1, \quad s, t \in I, z \in \mathbb{R}^6.$$

PROOF. Most of this is standard theory for ordinary differential equations, in particular, $Z(r, t, Z(t, s, z)) = Z(r, s, z)$ by uniqueness. Hence $Z(s, t, Z(t, s, z)) = z$, i.e., $Z^{-1}(s, t, \cdot) = Z(t, s, \cdot)$. In order to see that the flow is measure preserving, we rewrite the characteristic system in the form

$$\dot{z} = G(s, z), \quad G(s, x, v) := (v, F(s, x)).$$

The assertion then follows from the fact that

$$\frac{d}{ds} \det \frac{\partial Z}{\partial z}(s, t, z) = \operatorname{div}_z G(s, Z(s, t, z)) \det \frac{\partial Z}{\partial z}(s, t, z) = 0;$$

divergence-free vector fields induce measure preserving flows, a fact also known as Liouville's theorem. $\square$

The relation between the characteristic flow and the Vlasov equation is as follows.

LEMMA 1.3. *Under the assumptions of Lemma 1.2 the following holds:*

- (a) A function $f \in C^1(I \times \mathbb{R}^6)$ satisfies the Vlasov equation

$$\partial_t f + v \cdot \partial_x f + F(t, x) \cdot \partial_v f = 0 \quad (1.6)$$

iff it is constant along every solution of the characteristic system (1.5).

(b) For $\mathring{f} \in C^1(\mathbb{R}^6)$ the function

$$f(t, z) := \mathring{f}(Z(0, t, z)), \quad t \in I, z \in \mathbb{R}^6$$

is the unique solution of (1.6) in the space $C^1(I \times \mathbb{R}^6)$ with $f(0) = \mathring{f}$. If $\mathring{f}$ is non-negative then so is f ,

$$\text{supp } f(t) = Z(t, 0, \text{supp } \mathring{f}), \quad t \in I,$$

and for every $p \in [1, \infty]$,

$$\|f(t)\|_p = \|\mathring{f}\|_p, \quad t \in I.$$

PROOF. For a solution $z(s)$ of the characteristic system (1.5),

$$\frac{d}{ds} f(s, z(s)) = (\partial_t f + v \cdot \partial_x f + F \cdot \partial_v f)(s, z(s)).$$

This proves (a), since through each point (t, x, v) there passes a characteristic curve. The remaining assertions follow immediately with Lemma 1.2. $\square$

Before giving a rigorous proof of Theorem 1.1 an “exploratory” computation is instructive, which the experts of the trade would usually accept as a proof in itself. For this computation let f be a solution on some time interval $[0, T[$. What are the crucial quantities that must be controlled in order to control the solution? Assuming that $f(t)$ has compact support as will indeed be the case for the solution constructed below, let

$$P(t) := \sup\{|v| \mid (x, v) \in \text{supp } f(t)\}, \quad t \in [0, T[. \quad (1.7)$$

Since $\|f(t)\|_\infty$ and $\|f(t)\|_1$ are constant by Lemma 1.3,

$$\|\rho(t)\|_\infty \leq C P^3(t), \quad \|\rho(t)\|_1 = C,$$

so that by Lemma P1,

$$\|\partial_x U(t)\|_\infty \leq C P^2(t).$$

By the characteristic system,

$$P(t) \leq P(0) + C \int_0^t P^2(s) ds, \quad (1.8)$$

where C depends only on $\mathring{f}$ and changes its value from line to line. This estimate gives local-in-time control on P and the quantities which we estimated against P . In order to get a smooth solution we need to control derivatives, i.e., we need to go through another

Gronwall loop as above, but for the differentiated quantities. The x -derivative of ρ can be estimated against the x -derivative of f and hence of the characteristics; note that the support of the solution is now under control. If we differentiate the characteristic system with respect to initial data we get a Gronwall inequality for $\partial_x Z(s, t, x, v)$, involving $\partial_x^2 U$ so that $\partial_x Z(s, t, x, v)$ will be bounded by the exponential of the time integral of $\partial_x^2 U$. The crucial point is that in the estimate for the latter quantity in Lemma P1(b), $\partial_x \rho$ enters only logarithmically, and the whole chain of estimates leads to a linear Gronwall estimate for $\partial_x \rho$. Hence the derivatives are under control as long as the function P is.

It is useful to go through the arguments of the above exploratory computation in the form of a rigorous proof at least once in a mathematical lifetime, and here is your chance:

PROOF OF THEOREM 1.1. We fix an initial datum $\mathring{f} \in C_c^1(\mathbb{R}^6)$ with $\mathring{f} \geq 0$. For later use we also fix two constants $\mathring{R} > 0$ and $\mathring{P} > 0$ such that

$$\mathring{f}(x, v) = 0 \quad \text{for } |x| \geq \mathring{R} \text{ or } |v| \geq \mathring{P}.$$

We consider the following iterative scheme. The 0th iterate is defined by

$$f_0(t, z) := \mathring{f}(z), \quad t \geq 0, z \in \mathbb{R}^6.$$

If the n th iterate $f_n : [0, \infty[\times \mathbb{R}^6 \rightarrow [0, \infty[$ is already defined, we define

$$\rho_n := \rho_{f_n}, \quad U_n := U_{\rho_n}$$

on $[0, \infty[\times \mathbb{R}^3$, and we denote by

$$Z_n(s, t, z) = (X_n, V_n)(s, t, x, v)$$

the solution of the characteristic system

$$\dot{x} = v, \quad \dot{v} = -\partial_x U_n(s, x)$$

with $Z_n(t, t, z) = z$. Then

$$f_{n+1}(t, z) := \mathring{f}(Z_n(0, t, z)), \quad t \geq 0, z \in \mathbb{R}^6,$$

defines the next iterate. The idea of the proof is to show that these iterates converge on some time interval in a sufficiently strong sense and to identify the limit as the desired solution.

Step 1. Using Lemmae 1.2, 1.3, and Lemma P1 it is a simple proof by induction to see that the iterates are well defined and enjoy the following properties:

$$f_n \in C^1([0, \infty[\times \mathbb{R}^6), \quad \|f(t)\|_\infty = \|\mathring{f}\|_\infty, \quad \|f(t)\|_1 = \|\mathring{f}\|_1, \quad t \geq 0,$$

$$f_n(t, x, v) = 0 \quad \text{for } |v| \geq P_n(t) \text{ or } |x| \geq \mathring{R} + \int_0^t P_n(s) ds,$$

where

$$\begin{aligned} P_0(t) &:= \mathring{P}, & P_n(t) &:= \sup\{|V_{n-1}(s, 0, z)| \mid z \in \text{supp } \mathring{f}, 0 \leq s \leq t\}, \quad n \in \mathbb{N}, \\ \rho_n &\in C^1([0, \infty[\times \mathbb{R}^3), \\ \|\rho(t)\|_1 &= \|\mathring{f}\|_1, & \|\rho(t)\|_\infty &\leq \frac{4\pi}{3} \|\mathring{f}\|_\infty P_n^3(t), \quad t \geq 0, \\ \rho_n(t, x) &= 0 \quad \text{for } |x| \geq \mathring{R} + \int_0^t P_n(s) \, ds, \end{aligned}$$

and finally

$$\partial_x U_n \in C^1([0, \infty[\times \mathbb{R}^3), \quad \|\partial_x U_n(t)\|_\infty \leq C(\mathring{f}) P_n^2(t),$$

where by Lemma P1(b) with $p = 1$,

$$C(\mathring{f}) := 4 \cdot 3^{1/3} \pi^{4/3} \|\mathring{f}\|_1^{1/3} \|\mathring{f}\|_\infty^{2/3}. \quad (1.9)$$

Since this particular constant enters into the length of the interval on which the iterates converge, the information on which parameters it depends is important for the proof of the continuation criterion.

Step 2. Let $P : [0, \delta[\rightarrow]0, \infty[$ denote the maximal solution of the integral equation

$$P(t) = \mathring{P} + C(\mathring{f}) \int_0^t P^2(s) \, ds,$$

i.e.,

$$P(t) = \mathring{P}(1 - \mathring{P}C(\mathring{f})t)^{-1}, \quad 0 \leq t < \delta := (\mathring{P}C(\mathring{f}))^{-1};$$

without loss of generality $\mathring{f} \neq 0$. We claim that for every $n \in \mathbb{N}_0$ and $t \in [0, \delta[$ the estimate

$$P_n(t) \leq P(t)$$

holds. The assertion is obvious for $n = 0$. Assume it holds for some $n \in \mathbb{N}_0$. Then by Step 1,

$$\begin{aligned} |V_n(s, 0, z)| &\leq |v| + \int_0^s \|\partial_x U_n(\tau)\|_\infty \, d\tau \leq \mathring{P} + C(\mathring{f}) \int_0^s P_n^2(\tau) \, d\tau \\ &\leq \mathring{P} + C(\mathring{f}) \int_0^s P^2(\tau) \, d\tau = P(s) \end{aligned}$$

for any $0 \leq s \leq t < \delta$ and $z \in \text{supp } \mathring{f}$ so that the assertion follows by induction. On the interval $[0, \delta[$ the following estimates hold:

$$\|\rho_n(t)\|_\infty \leq \frac{4\pi}{3} \|\mathring{f}\|_\infty P^3(t), \quad \|\partial_x U_n(t)\|_\infty \leq C(\mathring{f}) P^2(t), \quad n \in \mathbb{N}_0.$$

We aim to show that the iterative scheme converges uniformly on any compact subinterval of $[0, \delta]$. Hence we fix $0 < \delta_0 < \delta$. In order to estimate terms like

$$\partial_x U_n(t, X_n) - \partial_x U_n(t, X_{n+1}),$$

a bound on $\partial_x^2 U_n$ is needed, uniformly in n .

Step 3. There exists some constant $C > 0$ depending on the initial datum and on δ_0 such that

$$\|\partial_x \rho_n(t)\|_\infty + \|\partial_x^2 U_n(t)\|_\infty \leq C, \quad t \in [0, \delta_0], n \in \mathbb{N}_0.$$

In the following proof of this assertion the constant C may change its value from line to line; it is only important that it does not depend on $t \in [0, \delta_0]$ or on $n \in \mathbb{N}_0$. First we note that

$$|\partial_x \rho_{n+1}(t, x)| \leq \int_{|v| \leq P(t)} |\partial_x [\dot{f}(Z_n(0, t, x, v))]| dv \leq C \|\partial_x Z_n(0, t, \cdot)\|_\infty.$$

We fix $x, v \in \mathbb{R}^3$ and $t \in [0, \delta_0]$ and write $(X_n, V_n)(s)$ instead of $(X_n, V_n)(s, t, x, v)$. If we differentiate the characteristic system defining Z_n with respect to x we obtain the estimates

$$|\partial_x \dot{X}_n(s)| \leq |\partial_x V(s)|, \quad |\partial_x \dot{V}_n(s)| \leq \|\partial_x^2 U_n(s)\|_\infty |\partial_x X_n(s)|.$$

If we integrate these estimates, observe that $\partial_x X_n(t) = \text{id}$, $\partial_x V_n(t) = 0$, and add the results we find that

$$\begin{aligned} & |\partial_x X_n(s)| + |\partial_x V_n(s)| \\ & \leq 1 + \int_s^t (1 + \|\partial_x^2 U_n(\tau)\|_\infty) (|\partial_x X_n(\tau)| + |\partial_x V_n(\tau)|) d\tau. \end{aligned}$$

By Gronwall's lemma,

$$|\partial_x X_n(s)| + |\partial_x V_n(s)| \leq \exp \int_0^t (1 + \|\partial_x^2 U_n(\tau)\|_\infty) d\tau,$$

and hence

$$\|\partial_x \rho_{n+1}(t)\|_\infty \leq C \exp \int_0^t \|\partial_x^2 U_n(\tau)\|_\infty d\tau, \quad 0 \leq s \leq t \leq \delta_0.$$

We insert the estimate on ρ_{n+1} from Step 2 and the above estimate on $\partial_x \rho_{n+1}$ into the second estimate for $\partial_x^2 U_{n+1}$ from Lemma P1(b) to find that

$$\|\partial_x^2 U_{n+1}(t)\|_\infty \leq C \left(1 + \int_0^t \|\partial_x^2 U_n(\tau)\|_\infty d\tau \right).$$

By induction,

$$\|\partial_x^2 U_n(t)\|_\infty \leq C e^{Ct}, \quad t \in [0, \delta_0], n \in \mathbb{N}_0,$$

if we increase C so that $\|\partial_x^2 U_0\|_\infty \leq C$, and the claim of Step 3 is established.

Step 4. We show that the sequence (f_n) converges to some function f , uniformly on $[0, \delta_0] \times \mathbb{R}^6$. Firstly, for $n \in \mathbb{N}$, and $t \in [0, \delta_0]$, $z \in \mathbb{R}^6$,

$$|f_{n+1}(t, z) - f_n(t, z)| \leq C |Z_n(0, t, z) - Z_{n-1}(0, t, z)|.$$

For $0 \leq s \leq t$ we have, suppressing the t and z arguments of the characteristics,

$$\begin{aligned} |X_n(s) - X_{n-1}(s)| &\leq \int_s^t |V_n(\tau) - V_{n-1}(\tau)| d\tau, \\ |V_n(s) - V_{n-1}(s)| &\leq \int_s^t [|\partial_x U_n(\tau, X_n(\tau)) - \partial_x U_{n-1}(\tau, X_n(\tau))| \\ &\quad + |\partial_x U_{n-1}(\tau, X_n(\tau)) - \partial_x U_{n-1}(\tau, X_{n-1}(\tau))|] d\tau \\ &\leq \int_s^t [\|\partial_x U_n(\tau) - \partial_x U_{n-1}(\tau)\|_\infty \\ &\quad + C |X_n(\tau) - X_{n-1}(\tau)|] d\tau. \end{aligned}$$

If we add these estimates and apply Gronwall's lemma we obtain the estimate

$$\begin{aligned} |Z_n(s) - Z_{n-1}(s)| &\leq C \int_0^t \|\partial_x U_n(\tau) - \partial_x U_{n-1}(\tau)\|_\infty d\tau \\ &\leq C \int_0^t \|\rho_n(\tau) - \rho_{n-1}(\tau)\|_\infty^{2/3} \|\rho_n(\tau) - \rho_{n-1}(\tau)\|_1^{1/3} d\tau \\ &\leq C \int_0^t \|\rho_n(\tau) - \rho_{n-1}(\tau)\|_\infty d\tau \\ &\leq C \int_0^t \|f_n(\tau) - f_{n-1}(\tau)\|_\infty d\tau; \end{aligned}$$

note that the support of both $\rho_n(t)$ and $f_n(t)$ is bounded, uniformly in n and $t \in [0, \delta_0]$. Summing up we obtain

$$\|f_{n+1}(t) - f_n(t)\|_\infty \leq C_* \int_0^t \|f_n(\tau) - f_{n-1}(\tau)\|_\infty d\tau,$$

and by induction,

$$\|f_{n+1}(t) - f_n(t)\|_\infty \leq C \frac{C_*^n t^n}{n!} \leq C \frac{C^n}{n!}, \quad n \in \mathbb{N}_0, 0 \leq t \leq \delta_0.$$

This implies that the sequence is uniformly Cauchy and converges uniformly on $[0, \delta_0] \times \mathbb{R}^6$ to some function $f \in C([0, \delta_0] \times \mathbb{R}^6)$. The limit has the following properties:

$$f(t, x, v) = 0 \quad \text{for } |v| \geq P(t) \text{ or } |x| \geq \mathring{R} + \int_0^t P(s) \, ds$$

and

$$\rho_n \rightarrow \rho := \rho_f, \quad U_n \rightarrow U := U_f$$

as $n \rightarrow \infty$, uniformly on $[0, \delta_0] \times \mathbb{R}^3$.

Step 5. In this step we show that the limiting function f has the regularity required of a solution to the Vlasov–Poisson system. Since

$$\|\partial_x U_n(t) - \partial_x U_m(t)\|_\infty \leq C \|\rho_n(t) - \rho_m(t)\|_\infty^{2/3} \|\rho_n(t) - \rho_m(t)\|_1^{1/3}$$

and

$$\begin{aligned} & \|\partial_x^2 U_n(t) - \partial_x^2 U_m(t)\|_\infty \\ & \leq C \left[\left(1 + \ln \frac{R}{d} \right) \|\rho_n(t) - \rho_m(t)\|_\infty \right. \\ & \quad \left. + d \|\partial_x \rho_n(t) - \partial_x \rho_m(t)\|_\infty + R^{-3} \|\rho_n(t) - \rho_m(t)\|_1 \right] \end{aligned}$$

for any $0 < d \leq R$ the sequences $(\partial_x U_n)$ and $(\partial_x^2 U_n)$ are uniformly Cauchy on $[0, \delta_0] \times \mathbb{R}^3$; notice that due to the compact support in x , uniformly in n , the L^1 -difference of the ρ 's can be estimated against the L^∞ -difference which converges to zero by the previous step, and while the L^∞ -difference of the derivatives of the ρ 's can according to Step 3 only be estimated by a uniform and not necessarily small constant, it has the factor d in front which can be chosen smaller than any prescribed ε . Hence

$$U, \partial_x U, \partial_x^2 U \in C([0, \delta_0] \times \mathbb{R}^3).$$

This in turn implies that

$$Z := \lim_{n \rightarrow \infty} Z_n \in C^1([0, \delta_0] \times [0, \delta_0] \times \mathbb{R}^6),$$

which is the characteristic flow induced by the limiting field $-\partial_x U$. Hence

$$f(t, z) = \lim_{n \rightarrow \infty} \mathring{f}(Z_n(0, t, z)) = \mathring{f}(Z(0, t, z)),$$

and $f \in C^1([0, \delta_0] \times \mathbb{R}^6)$ is a classical solution of the Vlasov–Poisson system. Since the arguments from Steps 3–5 hold on any compact subinterval of the interval $[0, \delta]$ this solution exists on the latter interval, and it is straight forward to verify the remaining properties

from the above definition of a classical solution such as the differentiability of ρ and U with respect to t .

Step 6. In order to show uniqueness we take two solutions f and g according to the definition with $f(0) = g(0)$, which both exist on some interval $[0, \delta]$. By (iii) in the definition of solution and Lemma 1.3 both $f(t)$ and $g(t)$ are supported in a compact set in $\mathbb{R}^6$ which can be chosen independent of $t \in [0, \delta]$. The estimates for the difference of two iterates $f_n - f_{n-1}$ can now be repeated for the difference $f - g$ to obtain the estimate

$$\|f(t) - g(t)\|_\infty \leq C \int_0^t \|f(s) - g(s)\|_\infty ds$$

on the interval $[0, \delta]$, and uniqueness follows.

Step 7. In order to prove the continuation criterion, let $f \in C^1([0, T[\times \mathbb{R}^6)$ be the maximally extended classical solution obtained above, and assume that

$$P^* := \sup\{|v| \mid (t, x, v) \in \text{supp } f\} < \infty,$$

but $T < \infty$. By Lemma 1.3,

$$\|f(t)\|_\infty = \|\dot{f}\|_\infty, \quad \|f(t)\|_1 = \|\dot{f}\|_1, \quad 0 \leq t < T.$$

The idea is to use the control of the length δ of the interval on which we constructed the solution in Steps 1–5 to show that, if we use the procedure above for the new initial value problem where we prescribe $f(\dot{t})$ as initial datum at time $t = \dot{t}$, we extend the solution beyond T if $\dot{t}$ is chosen sufficiently close to T . This is then the desired contradiction. To carry this out we notice first that $C(f(\dot{t})) = C(\dot{f})$, cf. (1.9). The maximal solution of the equation

$$P(t) = P^* + C(f(\dot{t})) \int_{\dot{t}}^t P^2(s) ds$$

exists on some interval $[\dot{t}, \dot{t} + \delta^*]$ the length δ^* of which is independent of $\dot{t}$. But since $f(\dot{t}, x, v) = 0$ for $|v| \geq P^*$ by definition of the latter quantity, the functions P_n will be bounded by P on this interval, and all the estimates from Steps 2–5 can be repeated on the interval $[\dot{t}, \dot{t} + \delta^*]$ so that our solution does exist there. If the a priori bound on ρ holds this gives an a priori bound on the field $-\partial_x U$ and hence on the quantity P^* as well, and the proof is complete. $\square$

Concluding remarks. (a) The above proof is essentially given in [7] although the result is not stated there.

(b) The analogous result, in particular the analogous continuation criterion is valid for the relativistic Vlasov–Maxwell system [36] or the Vlasov–Einstein system in the case of spherical symmetry and asymptotic flatness [100]; for these systems the control of the velocity support has to be replaced by control of the momentum support. Due to the nonlinear nature of the Einstein equations the continuation criterion is not valid for the Vlasov–Einstein system in general.

- (c) Uniqueness within weaker solution concepts is considered in [81,103,114].
 (d) Uniqueness is violated within the framework of measure-valued, weak solutions to the Vlasov–Poisson system in one space dimension [82], cf. also [83].

1.3. Spherically symmetric solutions

The estimates used in the proof of the local existence result are not strong enough to yield global existence. Indeed, it is a priori not clear that solutions should exist globally in time. In the plasma physics case one might argue that the particles repulse each other and hence the spatial density should remain bounded. But in the gravitational case the particles attract each other, and a gravitational collapse seems conceivable. To make these doubts more substantial we consider spherically symmetric solutions which by definition are invariant under simultaneous rotations of both x and v ,

$$f(t, x, v) = f(t, Ax, Av), \quad A \in \text{SO}(3).$$

If f is a solution, this transformation produces another one which by uniqueness coincides with f if the initial datum is spherically symmetric. Hence spherical symmetry is preserved by the Vlasov–Poisson system. For a spherically symmetric solution

$$\rho(t, x) = \rho(t, r), \quad U(t, x) = U(t, r), \quad r := |x|,$$

with some abuse of notation, and

$$\partial_r U(t, r) = \frac{4\pi\gamma}{r^2} \int_0^r \rho(t, s) s^2 ds, \quad \partial_x U(t, x) = \partial_r U(t, r) \frac{x}{r}. \quad (1.10)$$

Example of a “dust” solution which blows up. Let $\gamma = 1$. A likely candidate for a gravitational collapse is a spherically symmetric ensemble of particles which all move radially inward in such a way that they all arrive at the center at the same time. Hence let us consider a distribution function of the form

$$f(t, x, v) = \rho(t, x) \delta(v - u(t, x)),$$

where δ is the Dirac distribution. In such a distribution there is at each point in space only one particle velocity given by the velocity field $u(t, x) \in \mathbb{R}^3$, in other words, there is no velocity scattering. Notice that such an f is forbidden in our definition of solution, more importantly, it is not a distribution *function* on phase space, but we consider it anyway. Formally, such an f satisfies the Vlasov–Poisson system provided the spatial density ρ and the velocity field u , which now are the dynamical variables, satisfy the following special case of the Euler–Poisson system

$$\partial_t \rho + \text{div}(\rho u) = 0, \quad (1.11)$$

$$\partial_t u + (u \cdot \partial_x) u = -\partial_x U, \quad (1.12)$$

$$\Delta U = 4\pi\rho, \quad \lim_{|x| \rightarrow \infty} U(t, x) = 0.$$

This system represents a self-gravitating, ideal, compressible fluid with the strange equation of state that the pressure is zero, a situation which in astrophysics is sometimes referred to as *dust*. More information on how to pass from the Vlasov–Poisson system to the pressure-less Euler–Poisson system can be found in [20]. As initial datum to this system we take a homogeneous ball about the origin with all the particles at rest

$$\rho(0) = \frac{3}{4\pi} \mathbb{1}_{B_1}, \quad u(0) = 0.$$

For such data the following ansatz is reasonable:

$$\rho(t, x) = \frac{3}{4\pi} \frac{1}{R^3(t)} \mathbb{1}_{B_{R(t)}}(x), \quad u(t, x) = \frac{\dot{R}(t)}{R(t)} x,$$

i.e., we assume that the system retains the shape of a homogeneous ball, but the ball may contract (or expand). The radius $R: [0, T[\rightarrow]0, \infty[$ of the ball has to be determined such that a solution of the system above is obtained. It is straight forward to see that the continuity equation (1.11) is satisfied for $|x| \neq R(t)$; it holds in a weak sense everywhere. Moreover,

$$\begin{aligned} \partial_x U(t, x) &= R^{-3}(t)x, & \partial_t u(t, x) &= \left(\frac{\ddot{R}}{R} - \frac{\dot{R}^2}{R^2} \right)(t)x, \\ (u \cdot \partial_x)u &= \frac{\dot{R}^2}{R^2}(t)x. \end{aligned}$$

Hence Newton's law (1.12) is equivalent to the equation

$$\ddot{R} = -R^{-2} \tag{1.13}$$

which is nothing but the equation for the radial motion of a mass point in a central gravitational field. Our initial data translate into the condition

$$R(0) = 1, \quad \dot{R}(0) = 0.$$

The corresponding solution of (1.13) becomes 0 in finite time which means that all the mass of the solution of the pressure-less Euler–Poisson system collapses to a point in finite time.

This blow-up result for a system which does not belong to kinetic theory but seems to be closely related to the Vlasov–Poisson system might motivate one to look for a corresponding blow-up example for the latter. But in 1977, J. Batt proved the following result, which was the first global existence result for the Vlasov–Poisson system in three space dimensions – the result holds for both $\gamma = 1$ and $\gamma = -1$.

THEOREM 1.4. Let $\mathring{f} \in C_c^1(\mathbb{R}^6)$, $\mathring{f} \geq 0$ be spherically symmetric. Then there exists a constant $P_0 > 0$ such that for the corresponding classical solution of the Vlasov–Poisson system

$$f(t, x, v) = 0 \quad \text{for } |v| \geq P_0, 0 \leq t < T, x \in \mathbb{R}^3,$$

in particular, the solution is global in time, $T = \infty$. The constant P_0 depends only on $\|\mathring{f}\|_1$, $\|\mathring{f}\|_\infty$, and $\mathring{P}$, where $\mathring{f}(x, v) = 0$ for $|v| \geq \mathring{P}$.

PROOF. With $M := \|\mathring{f}\|_1 = \|f(t)\|_1$, formula (1.10) implies that

$$|\partial_x U(t, x)| \leq \frac{M}{r^2}, \quad r = |x|, t \geq 0.$$

On the other hand, by Lemma P1(b) with $p = 1$,

$$|\partial_x U(t, x)| \leq C \|\rho(t)\|_\infty^{2/3} \leq C P^2(t);$$

for technical reasons P is redefined to be nondecreasing:

$$P(t) := \sup\{|v| \mid (x, v) \in \text{supp } f(s), 0 \leq s \leq t\}.$$

Combining both estimates we find that

$$|\partial_x U(t, x)| \leq C \min\left\{\frac{1}{r^2}, P^2(t)\right\}.$$

Hence for any characteristic $(x(s), v(s))$ which starts in the support of $\mathring{f}$ we have for $i = 1, 2, 3$ and $0 \leq s \leq t < T$ the estimate

$$\begin{aligned} |\ddot{x}_i(s)| &\leq |\partial_{x_i} U(s, x(s))| \\ &\leq C^* \min\left\{\frac{1}{|x_i(s)|^2}, P^2(t)\right\}, \end{aligned}$$

where the constant C^* depends only on the L^1 and L^∞ -norms of $\mathring{f}$. Let $\xi := x_i$. Then $\xi \in C^2([0, t])$ with

$$|\ddot{\xi}(s)| \leq g(\xi(s)), \quad 0 \leq s \leq t,$$

where

$$g(r) := C^* \min\left\{\frac{1}{r^2}, P^2(t)\right\} \geq 0, \quad r \in \mathbb{R}.$$

If $\dot{\xi}(s) \neq 0$ on $]0, t[$, i.e., $\dot{\xi}$ does not change sign, it follows that

$$\begin{aligned} |\dot{\xi}(t) - \dot{\xi}(0)|^2 &\leq |\dot{\xi}(t) - \dot{\xi}(0)| |\dot{\xi}(t) + \dot{\xi}(0)| \\ &= |\dot{\xi}(t)^2 - \dot{\xi}(0)^2| = 2 \left| \int_0^t \dot{\xi}(s) \ddot{\xi}(s) \, ds \right| \\ &\leq 2 \int_0^t |\dot{\xi}(s)| g(\xi(s)) \, ds = 2 \int_{\xi([0, t])} g(r) \, dr \\ &\leq 2 \int g(r) \, dr = 8C^* P(t), \end{aligned}$$

and hence

$$|\dot{\xi}(t) - \dot{\xi}(0)| \leq 2\sqrt{2C^*} P^{1/2}(t).$$

If $\dot{\xi}(s) = 0$ for some $s \in]0, t[$ we define

$$s_- := \inf\{s \in]0, t[\mid \dot{\xi}(s) = 0\}, \quad s_+ := \sup\{s \in]0, t[\mid \dot{\xi}(s) = 0\}$$

so that $0 \leq s_- \leq s_+ \leq t$, $\dot{\xi}(s_-) = \dot{\xi}(s_+) = 0$, and the first case applies on the intervals $[0, s_-]$ and $[s_+, t]$. Hence

$$|\dot{\xi}(t) - \dot{\xi}(0)| \leq |\dot{\xi}(t) - \dot{\xi}(s_+)| + |\dot{\xi}(s_-) - \dot{\xi}(0)| \leq 4\sqrt{2C^*} P^{1/2}(t).$$

Since $\dot{\xi} = \dot{x}_i = v_i$, this implies that

$$P(t) \leq P(0) + 4\sqrt{6C^*} P^{1/2}(t), \quad t \in [0, T[.$$

The proof is complete. □

Given the blow-up example for the pressure-less Euler–Poisson system on the one hand and the global existence result for the spherically symmetric Vlasov–Poisson system on the other, the question arises whether there are similar semi-explicit solutions to the Vlasov–Poisson system and how they behave. A family of such examples has been constructed by Kurth [70].

Semiexplicit spherically symmetric solutions. It is easy to check that

$$f_0(x, v) := \frac{3}{4\pi^3} \begin{cases} (1 - |x|^2 - |v|^2 + |x \times v|^2)^{-1/2}, \\ \quad \text{where } (\cdot \cdot \cdot) > 0 \text{ and } |x \times v| < 1, \\ 0 & \text{else} \end{cases}$$

defines a time independent solution with spatial density and potential

$$\rho_0(x) = \frac{3}{4\pi} \mathbb{1}_{B_1}(x), \quad U_0(x) = \begin{cases} |x|^2/2 - 3/2, & |x| \leq 1, \\ -1/|x|, & |x| > 1. \end{cases}$$

Note that due to spherical symmetry the particle angular momentum $x \times v$ is preserved along characteristics, the particle energy $|v|^2/2 + U_0(x)$ is preserved because U_0 is time independent, and f_0 is a function of these invariants. The transformation

$$f(t, x, v) := f_0\left(\frac{x}{R(t)}, R(t)v - \dot{R}(t)x\right)$$

turns this steady state into a time dependent solution with spatial mass density

$$\rho(t) = \frac{3}{4\pi} \frac{1}{R^3(t)} \mathbb{1}_{B_{R(t)}},$$

provided the function $R = R(t)$ solves the differential equation

$$\ddot{R} - R^{-3} + R^{-2} = 0,$$

and $R(0) = 1$. Notice that in this example the spatial density is constant on a ball with a time dependent radius, like for the Euler–Poisson example stated above. Depending on $\alpha := \dot{R}(0)$ the solution behaves as follows:

- If $\alpha = 0$ then $R(t) = 1$, $t \in \mathbb{R}$, and we recover the steady state f_0 .
- If $0 < |\alpha| < 1$ then

$$R(t) = \frac{1 - \alpha \cos \phi(t)}{1 - \alpha^2},$$

where $\phi(t)$ is uniquely determined by

$$\begin{aligned} \phi(t) - \alpha \sin \phi(t) &= (1 - \alpha^2)^{3/2}(t - t_0), \\ t_0 &:= -(1 - \alpha^2)^{-3/2}(\phi_0 - \alpha \sin \phi_0), \quad \phi_0 := \arccos \alpha. \end{aligned}$$

The solution is time periodic with period $2\pi(1 - \alpha^2)^{-3/2}$.

- If $|\alpha| = 1$ then

$$R(t) = \frac{1 + \phi^2(t)}{2},$$

where $\phi(t)$ is uniquely determined by

$$\phi(t) + \frac{\phi^3(t)}{3} = 2\left(\alpha t + \frac{2}{3}\right).$$

The solution is global, but $R(t) \rightarrow \infty$ for $|t| \rightarrow \infty$, and R is strictly decreasing on $] -\infty, t_0]$ and strictly increasing on $[t_0, \infty[$ where $t_0 = -2/(3\alpha)$.

- If $|\alpha| > 1$ then

$$R(t) = \frac{|\alpha| \cosh \phi(t) - 1}{\alpha^2 - 1},$$

where $\phi(t)$ is uniquely determined by

$$\begin{aligned} \phi(t) - |\alpha| \sinh \phi(t) &= -(\alpha^2 - 1)^{3/2}(t - t_0), \\ t_0 &:= (\alpha^2 - 1)^{-3/2}(\phi_0 - |\alpha| \sinh \phi_0), \quad \cosh \phi_0 = \alpha, \operatorname{sgn} \phi_0 = \operatorname{sgn} \alpha. \end{aligned}$$

The solution is global with $R(t) \rightarrow \infty$ for $|t| \rightarrow \infty$, and R is strictly decreasing on $] -\infty, t_0]$ and strictly increasing on $[t_0, \infty[$.

Rigorously speaking, this example does not fit into our definition of solution, because f becomes singular at the boundary of its support, but the induced field allows for well defined characteristics, and f is constant along these. In no case does the solution blow up. To understand this difference to the dust example discussed above it should be observed that as opposed to the former there is velocity scattering in these solutions.

Concluding remarks. (a) The original proof of Theorem 1.4 given in [7] considered only the gravitational case $\gamma = 1$, which is the more difficult case anyway. It relied on a detailed analysis of the characteristic system, written in coordinates adapted to the symmetry: For a spherically symmetric solution, $f(t, x, v) = f(t, r, u, \alpha)$, where $r := |x|$, $u := |v|$, and α is the angle between x and v . It is easy to check that the modulus of angular momentum $ru \sin \alpha = |x \times v|$ is conserved along characteristics, and this fact was exploited in [7]. The above proof of Theorem 1.4 is due to Horst [58], where an analogous result is shown also for axially symmetric solutions which by definition are invariant under rotations about some fixed axis. To prove the latter result a priori bounds on the kinetic energy of the solution and on $\|\rho(t)\|_{5/3}$ were established first. These a priori bounds are discussed in Section 1.5, since they become essential for the global results in Section 1.6.

(b) Angular momentum invariants have proven useful in related situations. For example, in the plasma physics case global existence of classical solutions to the relativistic Vlasov–Poisson system has been shown for spherically symmetric and for axially symmetric initial data [29,34]. In the gravitational case blow-up occurs for that system, cf. Section 1.7.

(c) Under the assumption of spherical symmetry the relativistic Vlasov–Maxwell system reduces to the plasma physics case of the relativistic Vlasov–Poisson system, and global existence holds.

(d) The above Kurth solutions are to our knowledge the only time dependent solutions to the Vlasov–Poisson system, the behavior of which can be determined analytically. Notice that the boundary condition $\lim_{|x| \rightarrow \infty} U(t, x) = 0$ is part of our formulation of this system, i.e., we consider only isolated systems. Other semi-explicit solution families which do not satisfy this boundary condition are considered in [8]. For a cosmological interpretation of these solutions, which do not represent bounded particle ensembles, we refer to [25]. No Kurth type examples are known for the related systems.

1.4. Small data solutions

For a nonlinear evolution equation a natural question is whether sufficiently small initial data lead to solutions which decay and hence are global in time. This happens if the linear part of the equation has some dispersive property which is strong enough to dominate the nonlinearity as long as the solution is small. By no means all nonlinear PDEs have this property, but kinetic equations as a rule do. For the Vlasov–Poisson system this was established in [5]. In the following discussion of this result initial data are always taken from the set

$$\mathcal{D} := \left\{ \mathring{f} \in C_c^1(\mathbb{R}^6) \mid \mathring{f} \geq 0, \|\mathring{f}\|_\infty \leq 1, \|\partial_{(x,v)} \mathring{f}\|_\infty \leq 1, \right. \\ \left. \mathring{f}(x, v) = 0 \text{ for } |x| \geq \mathring{R} \text{ or } |v| \geq \mathring{P} \right\},$$

where $\mathring{R}, \mathring{P} > 0$ are arbitrary but fixed. Constants denoted by C may depend on these parameters and may change from line to line. The following theorem holds for both $\gamma = 1$ and $\gamma = -1$.

THEOREM 1.5. *There exists some $\delta > 0$ such that for any initial datum $\mathring{f} \in \mathcal{D}$ with $\|\mathring{f}\|_\infty < \delta$ the corresponding solution is global and satisfies the following decay estimates for $t > 0$:*

$$\|\rho(t)\|_\infty \leq Ct^{-3}, \quad \|\partial_x U(t)\|_\infty \leq Ct^{-2}, \quad \|\partial_x^2 U(t)\|_\infty \leq Ct^{-3} \ln(1+t).$$

The idea of the proof is as follows:

- If the field is zero, $\partial_x U = 0$, the free motion of the particles causes ρ to decay:

$$\begin{aligned} \rho(t, x) &= \int \mathring{f}(X(0, t, x, v), V(0, t, x, v)) dv = \int \mathring{f}(x - tv, v) dv \\ &= t^{-3} \int \mathring{f}\left(X, \frac{x - X}{t}\right) dX \leq Ct^{-3}. \end{aligned}$$

If the field is not zero but decays sufficiently fast this argument remains valid, i.e., the determinant of the matrix $\partial_v X(0, t, x, v)$ which comes up in the change of variables above grows like t^3 .

- By Lemma P1 a decay of ρ translates into a decay of the field.

If the decay of the field which is needed in the first step is asymptotically slower than the one resulting in the second step, then one can “bootstrap” this argument and obtain the decay estimates on the whole existence interval of the solution, which implies that the solution is global.

We approach the result through a series of lemmas, the first one being a local perturbation result about the trivial solution. It provides some finite time interval which can be made as long as desired and on which the solution exists and the field is sufficiently small to start the bootstrap argument outlined above.

LEMMA 1.6. *For any $\varepsilon > 0$ and $T > 0$ there exists some $\delta > 0$ such that every solution with initial datum $\mathring{f} \in \mathcal{D}$ satisfying $\|\mathring{f}\|_\infty < \delta$ exists on the interval $[0, T]$ and satisfies the estimate*

$$\|\partial_x U(t)\|_\infty + \|\partial_x^2 U(t)\|_\infty \leq \varepsilon, \quad t \in [0, T].$$

PROOF. By Step 2 of the proof of Theorem 1.1 the solution for any initial datum $\mathring{f} \in \mathcal{D}$ exists on the interval $[0, (C(\mathring{f})\mathring{P})^{-1}]$, where by (1.9),

$$C(\mathring{f}) = 4 \cdot 3^{1/3} \pi^{4/3} \|\mathring{f}\|_1^{1/3} \|\mathring{f}\|_\infty^{2/3} \leq C_0 \|\mathring{f}\|_\infty,$$

with C_0 depending only on the parameters $\mathring{R}$ and $\mathring{P}$. Hence if $\delta := (2C_0\mathring{P}T)^{-1}$, the solution exists on the prescribed time interval $[0, T]$, provided $\|\mathring{f}\|_\infty < \delta$. Let $\tilde{P} : [0, 2T[\rightarrow]0, \infty[$ denote the maximal solution of

$$\tilde{P}(t) = \mathring{P} + \frac{1}{2T\mathring{P}} \int_0^t \tilde{P}^2(s) ds,$$

a function which depends only on $\mathring{R}, \mathring{P}$ and T . If $\|\mathring{f}\|_\infty < \delta$ then $C(\mathring{f}) \leq (2T\mathring{P})^{-1}$, and hence $f(t, x, v) = 0$ for $|v| \geq \tilde{P}(t)$ and $t \in [0, T]$. This implies that

$$\|\rho(t)\|_\infty \leq \frac{4\pi}{3} \tilde{P}^3(T) \|\mathring{f}\|_\infty,$$

and we also have

$$\|\rho(t)\|_1 \leq \left(\frac{4\pi}{3}\right)^2 \mathring{R}^3 \mathring{P}^3 \|\mathring{f}\|_\infty$$

for all $t \in [0, T]$. After making δ smaller if necessary, Lemma P1(b) implies the desired estimates for $\partial_x U$ and $\partial_x^2 U$. For the latter quantity we have to go through the estimates in Step 3 of the proof of Theorem 1.1 applied to the solution instead of the iterates to find that $\|\partial_x \rho(t)\|_\infty < C$ where the constant depends only on $\mathring{R}$ and $\mathring{P}$. Then by Lemma P1(b),

$$\|\partial_x^2 U(t)\|_\infty \leq C(\delta + d + (1 - \ln d)\delta)$$

for any $0 < d < 1$, $t \in [0, T]$ and $\mathring{f} \in \mathcal{D}$ which $\|\mathring{f}\|_\infty < \delta$. The right-hand side can be made less than ε by first choosing d sufficiently small and then again making δ smaller if necessary. $\square$

The following decay condition on the field is the substitute for $\partial_x U$ to vanish identically in the first step of the bootstrap argument. Let $a > 0$ and $\alpha > 0$. A solution satisfies the *free streaming condition with parameter α* on the interval $[0, a]$ if the solution exists on $[0, a]$ and satisfies the estimates

$$\begin{cases} \|\partial_x U(t)\|_\infty \leq \alpha(1+t)^{-3/2}, \\ \|\partial_x^2 U(t)\|_\infty \leq \alpha(1+t)^{-5/2} \end{cases} \quad (\text{FS}\alpha)$$

there. The next lemma justifies this terminology: Under the assumption the long time asymptotics of certain quantities are like in the case where the field vanishes identically, provided the parameter α is chosen sufficiently small.

LEMMA 1.7. *If $\alpha > 0$ is small enough then any solution f with initial datum $\mathring{f} \in \mathcal{D}$, which satisfies the free streaming condition (FS α) on some interval $[0, a]$, has the following properties for all $t \in [0, a]$:*

- (a) $f(t, x, v) = 0$ for $|v| \geq \mathring{P} + 1$ and $x \in \mathbb{R}^3$;
- (b) $|\det \partial_v X(0, t, x, v)| \geq \frac{1}{2} t^3$ for $(x, v) \in \mathbb{R}^6$;
- (c) for $t > 0, x \in \mathbb{R}^3$ the mapping $X(0, t, x, \cdot) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is a C^1 -diffeomorphism;
- (d) $\|\partial_x \rho(t)\|_\infty \leq 4\pi(\mathring{P} + 1)^3$.

PROOF. Let $s \mapsto (x(s), v(s))$ be a characteristic with $|v(0)| \leq \mathring{P}$. Then for any $t \in [0, a]$ by (FS α),

$$|v(t)| \leq \mathring{P} + \int_0^t \|\partial_x U(s)\|_\infty ds \leq \mathring{P} + \alpha \int_0^t (1+s)^{-3/2} ds \leq \mathring{P} + 2\alpha,$$

which implies (a) if $\alpha \leq 1/2$. As to (b), we define for $0 \leq s \leq t \leq a$ and $(x, v) \in \mathbb{R}^6$ the function

$$\xi(s) := \partial_v X(s, t, x, v) - (s - t)\text{id}.$$

Clearly,

$$\ddot{\xi}(s) = -\partial_x^2 U(s, X(s, t, x, v)) \cdot \partial_v X(s, t, x, v), \quad \xi(t) = \dot{\xi}(t) = 0,$$

and by (FS α),

$$|\ddot{\xi}(s)| \leq \alpha(1+s)^{-5/2}(|\xi(s)| + (t-s)).$$

Upon integrating this inequality twice and switching the order of integration we obtain the estimate

$$\begin{aligned} |\xi(s)| &\leq \int_s^t \int_\tau^t |\ddot{\xi}(\sigma)| d\sigma d\tau = \int_s^t \int_s^\sigma |\ddot{\xi}(\sigma)| d\tau d\sigma \\ &\leq \alpha \int_s^t (1+\sigma)^{-3/2} (|\xi(\sigma)| + (t-\sigma)) d\sigma \\ &\leq 2\alpha(t-s) + \alpha \int_s^t (1+\sigma)^{-3/2} |\xi(\sigma)| d\sigma. \end{aligned}$$

By Gronwall's lemma,

$$|\xi(s)| \leq 2\alpha(t-s) \exp\left(\alpha \int_s^t (1+\sigma)^{-3/2} d\sigma\right) \leq 2\alpha e^{2\alpha}(t-s).$$

If we take $s = 0$, recall the definition of ξ , and divide by t we can rewrite this as

$$\left| \frac{1}{t} \partial_v X(0, t, x, v) + \text{id} \right| \leq 2\alpha e^{2\alpha}. \quad (1.14)$$

The assertion in (b) follows if $\alpha > 0$ is sufficiently small. In addition, by (1.14) the mapping considered in (c) is one-to-one

$$\begin{aligned} & |X(0, t, x, v) - X(0, t, x, \bar{v})| \\ &= \left| \int_0^1 \partial_v X(0, t, x, \tau v + (1 - \tau)\bar{v})(v - \bar{v}) \, d\tau \right| \\ &= \left| \int_0^1 [-t \text{id} + t \text{id} + \partial_v X(0, t, x, \tau v + (1 - \tau)\bar{v})](v - \bar{v}) \, d\tau \right| \\ &\geq t|v - \bar{v}| - 2\alpha e^{2\alpha} t|v - \bar{v}| \geq \frac{1}{2} t|v - \bar{v}| \end{aligned}$$

for $v, \bar{v} \in \mathbb{R}^3$, $x \in \mathbb{R}^3$, and $t \in]0, a]$; for the last estimate α is again chosen smaller if necessary. Hence the mapping $X(0, t, x, \cdot)$ is a C^1 -diffeomorphism onto its range which is an open set. Assume that it were not onto $\mathbb{R}^3$. Then the range $X(0, t, x, \mathbb{R}^3)$ has a boundary point x_0 which is not an image point. Choose a sequence $(v_n) \subset \mathbb{R}^3$ such that $X(0, t, x, v_n) \rightarrow x_0$. By the previous estimate, $v_n \rightarrow v_0$ converges, and by continuity, $x_0 = X(0, t, x, v_0)$ is an image point. This is a contradiction, and the assertion in (c) is established. As to (d), clearly

$$\|\partial_x \rho(t)\|_\infty \leq \frac{4\pi}{3} (\mathring{P} + 1)^3 \|\partial_x f(t)\|_\infty$$

and

$$|\partial_x f(t, z)| \leq \|\partial_z \mathring{f}\|_\infty (|\partial_x X(0, t, z)| + |\partial_x V(0, t, z)|).$$

By definition of the initial data set $\mathcal{D}$, $\|\partial_z \mathring{f}\|_\infty \leq 1$, so it remains to estimate the derivatives of the characteristics. Proceeding as above we define

$$\xi(s) := \partial_x X(s, t, x, v) - \text{id}$$

so that

$$|\ddot{\xi}(s)| \leq \alpha(1+s)^{-5/2} (|\xi(s)| + 1), \quad \xi(t) = \dot{\xi}(t) = 0.$$

The resulting Gronwall estimate yields, for α sufficiently small,

$$|\xi(s)| \leq 2\alpha e^{2\alpha} \leq 1,$$

and

$$|\dot{\xi}(s)| \leq \int_s^t |\ddot{\xi}(\tau)| d\tau \leq \alpha \int_s^t (1+\tau)^{-5/2} (|\xi(\tau)| + 1) d\tau \leq 1.$$

Since $\partial_x V(0, t, x, v) = \dot{\xi}(0)$, we have shown that

$$|\partial_x X(0, t, z)| + |\partial_x V(0, t, z)| \leq 3,$$

and the proof is complete. $\square$

After these preparations we are ready to prove Theorem 1.5.

PROOF OF THEOREM 1.5. We start by fixing some $\alpha > 0$ sufficiently small for all the assertions of Lemma 1.7 to hold, and we consider some interval $[0, a]$ with $a > 1$ on which $(FS\alpha)$ holds for some solution f with initial datum $\mathring{f} \in \mathcal{D}$. For $t \in]0, a]$ and $x \in \mathbb{R}^3$ the change of variables $v \mapsto X = X(0, t, x, v)$ and Lemma 1.7(b),(c) imply that

$$\begin{aligned} \rho(t, x) &= \int \mathring{f}(X(0, t, x, v), V(0, t, x, v)) dv \\ &= \int \mathring{f}(X, V(0, t, x, v(X))) |\det \partial_v X^{-1}(0, t, x, v(X))| dX \\ &\leq \frac{8\pi}{3} \mathring{R}^3 \|\mathring{f}\|_\infty t^{-3}; \end{aligned}$$

$v(X)$ denotes the inverse of the change of variables. Hence by Lemma 1.7(d),

$$\|\rho(t)\|_\infty \leq C_1 t^{-3}, \quad \|\partial_x \rho(t)\|_\infty \leq C_1, \quad t \in [0, a],$$

where the constant C_1 depends only on $\mathring{R}$ and $\mathring{P}$. By Lemma P1,

$$\|\partial_x U(t)\|_\infty \leq 3(2\pi)^{2/3} \|\mathring{f}\|_1^{1/3} C_1^{2/3} t^{-2} \leq C_2 t^{-2},$$

and for $t \in [1, a]$ with $R = t$ and $d = t^{-3} \leq R$,

$$\|\partial_x^2 U(t)\|_\infty \leq C[t^{-3} + t^{-3} + t^{-3} \ln t^4] \leq C_2(1 + \ln t)t^{-3},$$

where again the constant C_2 depends only on $\mathring{R}$ and $\mathring{P}$. We fix some time $T_0 > 1$ such that for all $t \geq T_0$,

$$C_2 t^{-2} \leq \frac{\alpha}{2} (1+t)^{-3/2}, \quad C_2(1 + \ln t)t^{-3} \leq \frac{\alpha}{2} (1+t)^{-5/2}$$

which means that the decay obtained as output in the above estimates is stronger than the one in the free streaming condition (FS α). Lemma 1.6 provides $\delta > 0$ such that any solution launched by an initial datum $\mathring{f} \in \mathcal{D}$ with $\|\mathring{f}\|_\infty < \delta$ exists on the maximal existence interval $[0, T[$ with $T > T_0$, and

$$\|\partial_x U(t)\|_\infty + \|\partial_x^2 U(t)\|_\infty < \frac{\alpha}{2}(1 + T_0)^{-5/2}, \quad t \in [0, T_0].$$

By continuity the free streaming condition (FS α) holds on some interval $[0, T^*[$ with $T^* \in]T_0, T]$, and we choose T^* maximal with this property. On $[T_0, T^*[$,

$$\begin{aligned} \|\partial_x U(t)\|_\infty &\leq C_2 t^{-2} \leq \frac{\alpha}{2}(1 + t)^{-3/2}, \\ \|\partial_x^2 U(t)\|_\infty &\leq C_2(1 + \ln t)t^{-3} \leq \frac{\alpha}{2}(1 + t)^{-5/2}, \end{aligned}$$

which implies that $T^* = T$, and by Lemma 1.7(a) and the continuation criterion from Theorem 1.1, $T = \infty$. $\square$

Concluding remarks. (a) Lemma 1.6 is a special case of the fact that solutions depend continuously on initial data, cf. [90], Theorem 1.

(b) Global existence for small initial data was established for the relativistic Vlasov–Maxwell system in [37], and these techniques lead to analogous results for nearly neutral and nearly spherically symmetric data [30,90]. Similar techniques have been employed for the spherically symmetric, asymptotically flat Vlasov–Einstein system [100].

1.5. Conservation laws and a priori bounds

Conservation laws represent physically relevant properties of the system, and they lead to a priori bounds on the solutions used for the global existence result. As a matter of fact we have already stated and used one such conservation law, namely conservation of phase space volume: The characteristic flow of the Vlasov equation is measure preserving, cf. Lemma 1.2, and this leads to the a priori bounds

$$\|f(t)\|_p = \|\mathring{f}\|_p, \quad p \in [1, \infty],$$

as long as the solution exists. For $p = 1$ this is conservation of mass

$$\int \int f(t, x, v) dv dx = \int \rho(t, x) dx = M,$$

which can also be viewed as a consequence of the local mass conservation law

$$\partial_t \rho + \operatorname{div} j = 0, \tag{1.15}$$

where the mass current j is defined as

$$j(t, x) := \int v f(t, x, v) dv.$$

Equation (1.15) follows by integrating the Vlasov equation with respect to v and observing that the total v -divergence $\partial_x U \cdot \partial_v f = \operatorname{div}_v(f \partial_x U)$ vanishes upon integration. Conservation of phase space volume follows from the Vlasov equation alone, regardless of the field equation to which it is coupled. The resulting a priori bounds are much too weak to gain global existence, since in particular only a bound on the L^1 -norm of ρ results.

However, the system is also conservative. There is no dissipative mechanism in the system, and hence energy is conserved. It is a straightforward computation to see that for a classical solution the total energy

$$\frac{1}{2} \iint |v|^2 f(t, x, v) dv dx - \frac{\gamma}{8\pi} \int |\partial_x U(t, x)|^2 dx \quad (1.16)$$

is constant as long as the solution exists. There is however an immediate problem: In the gravitational case $\gamma = 1$ the energy does not have a definite sign, and hence it is conceivable that the individual terms in (1.16), kinetic and potential energy, become unbounded in finite time while the sum remains constant. This does indeed happen for solutions of the N -body problem when two bodies collide, and it also happens for the counterexample to global existence for the pressure-less Euler–Poisson system in Section 1.3. If we consider the plasma physics case both kinetic and potential energy are obviously bounded. But as we will see shortly the same is true also in the gravitational case, which may come as a surprise.

Since conservation of energy plays a vital role in the stability analysis in the second part of this treatise, our presentation in the rest of the present section is a bit more general than necessary for the existence problem. The kinetic and the potential energy of a state $f \in L^1_+(\mathbb{R}^6)$ are defined as

$$E_{\text{kin}}(f) := \frac{1}{2} \iint |v|^2 f(x, v) dv dx,$$

$$E_{\text{pot}}(f) := -\frac{\gamma}{8\pi} \int |\nabla U_f(x)|^2 dx = \frac{1}{2} \int U_f(x) \rho_f(x) dx.$$

The spatial density ρ_f is bounded in an appropriate norm by the kinetic energy $E_{\text{kin}}(f)$. The reason is that the kinetic energy is a second-order moment in velocity of f , while ρ_f is a zeroth order moment. For later purposes we prove a more general result than needed right now.

LEMMA 1.8. *For $k \geq 0$ we denote the k th order moment density and the k th order moment in velocity of a nonnegative, measurable function $f : \mathbb{R}^6 \rightarrow [0, \infty[$ by*

$$m_k(f)(x) := \int |v|^k f(x, v) dv$$

and

$$M_k(f) := \int m_k(f)(x) dx = \iint |v|^k f(x, v) dv dx.$$

Let $1 \leq p, q \leq \infty$ with $1/p + 1/q = 1$, $0 \leq k' \leq k < \infty$, and

$$r := \frac{k + 3/q}{k' + 3/q + (k - k')/p}.$$

If $f \in L_+^p(\mathbb{R}^6)$ with $M_k(f) < \infty$ then $m_{k'}(f) \in L^r(\mathbb{R}^3)$ and

$$\|m_{k'}(f)\|_r \leq c \|f\|_p^{(k-k')/(k+3/q)} M_k(f)^{(k'+3/q)/(k+3/q)}$$

where $c = c(k, k', p) > 0$.

PROOF. We split the v -integral defining $m_{k'}(f)$ into small and large v 's and optimize with respect to the splitting parameter, more precisely, for any $R > 0$,

$$\begin{aligned} m_{k'}(f)(x) &\leq \int_{|v| \leq R} |v|^{k'} f(x, v) dv + \int_{|v| > R} |v|^{k'} f(x, v) dv \\ &\leq \|f(x, \cdot)\|_p \left(\int_{|v| \leq R} |v|^{k'q} dv \right)^{1/q} + R^{k'-k} \int |v|^k f(x, v) dv \\ &\leq c \|f(x, \cdot)\|_p R^{k'+3/q} + R^{k'-k} m_k(f)(x), \end{aligned}$$

where we used Hölder's inequality. Let

$$R := \left[\frac{m_k(f)(x)}{\|f(x, \cdot)\|_p} \right]^{1/(k+3/q)},$$

which up to a constant is the choice which minimizes the right-hand side as a function of $R > 0$. Then

$$m_{k'}(f)(x) \leq c (\|f(x, \cdot)\|_p)^{(k-k')/(k+3/q)} (m_k(f)(x))^{(k'+3/q)/(k+3/q)}.$$

If we take this estimate to the power r and integrate in x we can by the definition of r again apply Hölder's inequality, and the assertion follows. $\square$

Together with Lemma P2 the potential energy can be estimated in terms of the kinetic energy in such a way that by conservation of energy both terms individually remain bounded along classical solutions also in the case $\gamma = 1$. This was first observed by Horst [57].

PROPOSITION 1.9. *Let f be a classical solution of the Vlasov–Poisson system on the time interval $[0, T[$ with induced spatial density ρ . Then for all $t \in [0, T[$,*

$$E_{\text{kin}}(f(t)), \quad |E_{\text{pot}}(f(t))|, \quad \|\rho(t)\|_{5/3} \leq C,$$

where the constant depends only on the initial datum $f(0) = \hat{f}$, more precisely on its L^1 and L^∞ -norms and its kinetic energy.

PROOF. If $\gamma = -1$ the kinetic and potential energy are both nonnegative and hence bounded by conservation of energy. The bound on ρ follows by Lemma 1.8 with $k = 2$, $k' = 0$, $p = \infty$, $q = 1$, $r = 5/3$:

$$\|\rho(t)\|_{5/3} \leq c \|f(t)\|_\infty^{2/5} E_{\text{kin}}(f(t))^{3/5}.$$

If $\gamma = 1$ we use Lemma P2(b) and Lemma 1.8 with $k = 2$, $k' = 0$, $p = 9/7$, $r = 6/5$ to obtain

$$|E_{\text{pot}}(f(t))| \leq c \|\rho(t)\|_{6/5}^2 \leq c \|f(t)\|_{9/7}^{3/2} E_{\text{kin}}(f(t))^{1/2} = C E_{\text{kin}}(f(t))^{1/2},$$

where the constant C has the claimed dependence. By conservation of energy

$$E_{\text{kin}}(f(t)) - C E_{\text{kin}}(f(t))^{1/2} \leq E_{\text{kin}}(f(t)) + E_{\text{pot}}(f(t)) \leq E_{\text{kin}}(\hat{f}),$$

which implies the bound on the kinetic energy also for the case $\gamma = 1$. □

With these additional bounds at hand one may hope to improve the estimates in the local existence result in such a way that global existence follows. Indeed, by Lemma P1(b) with $p = 5/3$ and Proposition 1.9,

$$\|\partial_x U(t)\|_\infty \leq C \|\rho(t)\|_\infty^{4/9} \leq C P^{4/3}(t) \tag{1.17}$$

with P as defined in (1.7). Hence

$$P(t) \leq P(0) + \int_0^t \|\partial_x U(s)\|_\infty \, ds \leq P(0) + C \int_0^t P^{4/3}(s) \, ds. \tag{1.18}$$

This certainly is an improvement compared to our first attempt at bounding P , cf. equation (1.8), but the improvement is not sufficient to yield a global bound.

One way to improve this argument is to observe that an a priori bound on a higher-order L^p -norm of $\rho(t)$ allows for a smaller power of the L^∞ -norm of $\rho(t)$ in the estimate (1.17) and thus for a smaller power of $P(s)$ in the Gronwall inequality (1.18). In the estimate (1.17) we would need an exponent less or equal to $1/3$ on $\|\rho(t)\|_\infty$ to obtain a Gronwall estimate on P leading to a global bound. If we compare this to Lemma P1(b) and use Lemma 1.8 with $p = \infty$, $k = 3$, $k' = 0$ we obtain a less demanding continuation criterion which we note for later use.

PROPOSITION 1.10. *If for a local solution f on its maximal existence interval $[0, T[$ the quantity $\|\rho(t)\|_2$ or $M_3(t)$ is bounded, then the solution is global.*

Concluding remarks. (a) The a priori bounds in Proposition 1.9 together with compactness properties of the solution operator to the Poisson equation can be used to prove the existence of global weak solutions for the Vlasov–Poisson system [4,61]. These solutions are not known to be unique nor are they known to satisfy the above conservation laws.

(b) For the relativistic Vlasov–Poisson system the kinetic energy

$$\iint \sqrt{1 + |v|^2} f(t, x, v) \, dv \, dx$$

is of lower order in v than in the nonrelativistic case, and the potential energy turns out to be of the same order in the sense of the above estimates. Indeed, in the gravitational case the a priori bounds from Proposition 1.9 do not hold and solutions can blow up, cf. Theorem 1.17. For the plasma physics case the bound on the kinetic energy yields only a bound on $\|\rho(t)\|_{4/3}$, and these a priori bounds are then too weak for the proofs of global existence in the next section to extend to the relativistic case.

1.6. Global existence for general data

The aim of this section is to prove the following theorem.

THEOREM 1.11. *Any nonnegative initial datum $\mathring{f} \in C_c^1(\mathbb{R}^6)$ launches a global classical solution of the Vlasov–Poisson system.*

Let $[0, T[$ be the right maximal existence interval of the local solution provided by Theorem 1.1; all the arguments apply also when going backward in time. For technical reasons we redefine the quantity $P(t)$ and make it nondecreasing

$$P(t) := \max\{|v| \mid (x, v) \in \text{supp } f(s), 0 \leq s \leq t\}.$$

We need to show that this function is bounded on bounded time intervals. By Proposition 1.10 it also suffices to bound a sufficiently high-order moment in v . This is the approach followed by Lions and Perthame [80]. The approach followed by Pfaffelmoser [89] is to fix a characteristic $(X, V)(t)$ along which the increase in velocity

$$|V(t) - V(t - \Delta)| \leq \int_{t-\Delta}^t \iint \frac{f(s, y, w)}{|y - X(s)|^2} \, dw \, dy \, ds \quad (1.19)$$

during the time interval $[t - \Delta, t]$ is estimated. In the Gronwall argument leading to (1.18) we first split x -space to obtain the estimate (1.17) and then split v -space to obtain the estimate for ρ in Proposition 1.9. Pfaffelmoser's idea is that instead of doing one after the other one should split (x, v) -space in (1.19) into suitably chosen sets. Since this approach is

more elementary and gives better estimates on the possible growth of the solution, we discuss it first, following a greatly simplified version due to Schaeffer [105,106]. The Lions–Perthame approach, which has the greater potential to generalize to related situations, is presented second.

1.6.1. The Pfaffmoser–Schaeffer proof. Let us single out one particle in our distribution, the increase in velocity of which we want to control over a certain time interval. Mathematically speaking, we fix a characteristic $(X, V)(t)$ with $(X, V)(0) \in \text{supp } f^\circ$, and we take $0 \leq \Delta \leq t < T$. After the change of variables

$$y = X(s, t, x, v), \quad w = V(s, t, x, v), \quad (1.20)$$

equation (1.19) takes the form

$$|V(t) - V(t - \Delta)| \leq \int_{t-\Delta}^t \iint \frac{f(t, x, v)}{|X(s, t, x, v) - X(s)|^2} dv dx ds, \quad (1.21)$$

because f is constant along the volume preserving characteristic flow. For parameters $0 < p \leq P(t)$ and $r > 0$, which will be specified later, we split the domain of integration in (1.21) into the following sets:

$$\begin{aligned} M_g &:= \{(s, x, v) \in [t - \Delta, t] \times \mathbb{R}^6 \mid |v| \leq p \vee |v - V(t)| \leq p\}, \\ M_b &:= \{(s, x, v) \in [t - \Delta, t] \times \mathbb{R}^6 \mid |v| > p \wedge |v - V(t)| > p \\ &\quad \wedge [|X(s, t, x, v) - X(s)| \leq r|v|^{-3} \\ &\quad \vee |X(s, t, x, v) - X(s)| \leq r|v - V(t)|^{-3}]\}, \\ M_u &:= \{(s, x, v) \in [t - \Delta, t] \times \mathbb{R}^6 \mid |v| > p \wedge |v - V(t)| > p \\ &\quad \wedge |X(s, t, x, v) - X(s)| > r|v|^{-3} \\ &\quad \wedge |X(s, t, x, v) - X(s)| > r|v - V(t)|^{-3}\}. \end{aligned}$$

The logic behind the names of these sets is as follows. In the set M_g velocities are bounded, either with respect to our frame of reference or with respect to the one particle which we singled out. Hence M_g is the *good* set – we know how to proceed if the velocities are bounded. The set M_b is the *bad* set, since here velocities are large, and in addition the particle whose contribution to the integral in (1.21) we are computing is close in space to the singled out particle, i.e., the singularity of the Newton force is strong. Notice however that the latter type of badness is coupled with the former via the condition $|X(s, t, x, v) - X(s)| \leq r|v|^{-3}$. Both M_g and M_b are going to be estimated in a straight forward manner, while on the set M_u the time integral in (1.21) will be exploited in a crucial way to bound its contribution in terms of the kinetic energy. It is the *ugly* set although the ideas involved in its estimate are beautiful.

To estimate the contribution of each of these sets to the integral in (1.21) the length of the time interval $[t - \Delta, t]$ is chosen in such a way that velocities do not change very much on that interval. Recall that by (1.17),

$$\|\partial_x U(t)\|_\infty \leq C^* P(t)^{4/3}, \quad t \in [0, T[,$$

for some $C^* > 0$ so if

$$\Delta := \min \left\{ t, \frac{p}{4C^* P(t)^{4/3}} \right\} \quad (1.22)$$

then

$$|V(s, t, x, v) - v| \leq \Delta C^* P(t)^{4/3} \leq \frac{1}{4} p, \quad s \in [t - \Delta, t], x, v \in \mathbb{R}^3. \quad (1.23)$$

The contribution of the good set M_g . For $(s, x, v) \in M_g$ by (1.20) and (1.23),

$$|w| < 2p \vee |w - V(s)| < 2p.$$

Hence the change of variables (1.20) implies the estimate

$$\int_{M_g} \frac{f(t, x, v)}{|X(s, t, x, v) - X(s)|^2} dv dx ds \leq \int_{t-\Delta}^t \int \frac{\tilde{\rho}(s, y)}{|y - X(s)|^2} dy ds,$$

where

$$\tilde{\rho}(s, y) := \int_{|w| < 2p \vee |w - V(s)| < 2p} f(s, y, w) dw \leq Cp^3,$$

and by Proposition 1.9,

$$\|\tilde{\rho}(s)\|_{5/3} \leq \|\rho(s)\|_{5/3} \leq C.$$

Therefore, by the estimate (1.17),

$$\int_{M_g} \frac{f(t, x, v)}{|X(s, t, x, v) - X(s)|^2} dv dx ds \leq Cp^{4/3} \Delta. \quad (1.24)$$

The contribution of the bad set M_b . For $(s, x, v) \in M_b$ by (1.20) and (1.23),

$$\begin{aligned} \frac{1}{2} p < |w| < 2|v| \wedge \frac{1}{2} p < |w - V(s)| < 2|v - V(t)| \\ \wedge [|y - X(s)| < 8r|w|^{-3} \vee |y - X(s)| < 8r|w - V(s)|^{-3}]. \end{aligned}$$

On the other hand, $|w| \leq P(t)$ and $|w - V(s)| \leq 2P(t)$ for $w \in \text{supp } f(s, y, \cdot)$, $0 \leq s \leq t$. Thus by (1.20) and since $\|f(s)\|_\infty = \|\dot{f}\|_\infty$,

$$\begin{aligned}
 & \int_{M_b} \frac{f(t, x, v)}{|X(s, t, x, v) - X(s)|^2} dv dx ds \\
 & \leq \int_{t-\Delta}^t \int_{\frac{1}{2}p < |w| \leq P(t)} \int_{|y-X(s)| < 8r|w|^{-3}} \frac{f(s, y, w)}{|y - X(s)|^2} dy dw ds \\
 & \quad + \int_{t-\Delta}^t \int_{\frac{1}{2}p < |w-V(s)| \leq 2P(t)} \int_{|y-X(s)| < 8r|w-V(s)|^{-3}} \frac{f(s, y, w)}{|y - X(s)|^2} dy dw ds \\
 & \leq Cr \ln \frac{4P(t)}{p} \Delta. \tag{1.25}
 \end{aligned}$$

The contribution of the ugly set M_u . The main idea in estimating the contribution of the set M_u is to integrate with respect to time first, using the fact that on M_u the distance of $X(s, t, x, v)$ from $X(s)$ can be bounded from below linearly in time. Let $(x, v) \in \mathbb{R}^6$ with $|v - V(t)| > p$ and define

$$d(s) := X(s, t, x, v) - X(s), \quad s \in [t - \Delta, t].$$

We Taylor-expand this difference to first order around a point $s_0 \in [t - \Delta, t]$ where the difference is minimal

$$|d(s_0)| = \min\{|d(s)| \mid t - \Delta \leq s \leq t\}.$$

To this end, we define

$$\bar{d}(s) := d(s_0) + (s - s_0)\dot{d}(s_0), \quad s \in [t - \Delta, t].$$

Then

$$d(s_0) = \bar{d}(s_0), \quad \dot{d}(s_0) = \dot{\bar{d}}(s_0),$$

and

$$|\ddot{d}(s) - \ddot{\bar{d}}(s)| = |\dot{V}(s, t, x, v) - \dot{V}(s)| \leq 2\|\partial_x U(s)\|_\infty \leq 2C^*P(t)^{4/3}.$$

Hence

$$\begin{aligned}
 |d(s) - \bar{d}(s)| & \leq C^*P(t)^{4/3}(s - s_0)^2 \leq C^*P(t)^{4/3}\Delta|s - s_0| \\
 & \leq \frac{1}{4}p|s - s_0| < \frac{1}{4}|v - V(t)||s - s_0|. \tag{1.26}
 \end{aligned}$$

On the other hand, by (1.23),

$$|\dot{d}(s_0)| = |V(s_0, t, x, v) - V(s_0)| \geq |v - V(t)| - \frac{1}{2}p > \frac{1}{2}|v - V(t)|,$$

and by the definition of s_0 , distinguishing the cases $s_0 = t - \Delta$, $s_0 \in]t - \Delta, t[$, and $s_0 = t$,

$$(s - s_0) d(s_0) \cdot \dot{d}(s_0) \geq 0.$$

Hence for all $s \in [t - \Delta, t]$ the estimate

$$|\bar{d}(s)|^2 \geq \frac{1}{4}|v - V(t)|^2 |s - s_0|^2$$

holds. Combining this with (1.26) finally implies that the estimate

$$|d(s)| \geq \frac{1}{4}|v - V(t)| |s - s_0| \quad (1.27)$$

holds for all $s \in [t - \Delta, t]$ and $(x, v) \in \mathbb{R}^6$ with $|v - V(t)| > p$. To exploit this we define auxiliary functions

$$\sigma_1(\xi) := \begin{cases} \xi^{-2}, & \xi > r|v|^{-3}, \\ (r|v|^{-3})^{-2}, & \xi \leq r|v|^{-3}, \end{cases}$$

and

$$\sigma_2(\xi) := \begin{cases} \xi^{-2}, & \xi > r|v - V(t)|^{-3}, \\ (r|v - V(t)|^{-3})^{-2}, & \xi \leq r|v - V(t)|^{-3}. \end{cases}$$

The definition of M_u , the fact that the functions σ_i are nonincreasing and the estimate (1.27) imply that

$$|d(s)|^{-2} \mathbb{1}_{M_u}(s, x, v) \leq \sigma_i(|d(s)|) \leq \sigma_i\left(\frac{1}{4}|v - V(t)| |s - s_0|\right)$$

for $i = 1, 2$ and $s \in [t - \Delta, t]$. Hence we can estimate the time integral in the contribution of M_u in the following way:

$$\begin{aligned} \int_{t-\Delta}^t |d(s)|^{-2} \mathbb{1}_{M_u}(s, x, v) ds &\leq 8|v - V(t)|^{-1} \int_0^\infty \sigma_i(\xi) d\xi \\ &= 16|v - V(t)|^{-1} \begin{cases} r^{-1}|v|^3, & i = 1, \\ r^{-1}|v - V(t)|^3, & i = 2, \end{cases} \end{aligned}$$

and since this estimate holds for both $i = 1$ and $i = 2$,

$$\begin{aligned} & \int_{t-\Delta}^t |d(s)|^{-2} \mathbb{1}_{M_u}(s, x, v) \, ds \\ & \leq 16r^{-1} |v - V(t)|^{-1} \min\{|v|^3, |v - V(t)|^3\} \leq 16r^{-1} |v|^2. \end{aligned}$$

Therefore,

$$\begin{aligned} & \int_{M_u} \frac{f(t, x, v) \, dv \, dx \, ds}{|X(s, t, x, v) - X(s)|^2} \\ & \leq \iint f(t, x, v) \int_{t-\Delta}^t |d(s)|^{-2} \mathbb{1}_{M_u}(s, x, v) \, ds \, dv \, dx \\ & \leq Cr^{-1} \iint |v|^2 f(t, x, v) \, dv \, dx \\ & \leq Cr^{-1}, \end{aligned} \tag{1.28}$$

since according to Proposition 1.9 the kinetic energy is bounded.

Adding up the estimates (1.24), (1.25), (1.28) we arrive at the following control on the increase in velocity along the characteristic which we singled out

$$\begin{aligned} |V(t) - V(t - \Delta)| & \leq C \left(p^{4/3} + r \ln \frac{4P(t)}{p} + r^{-1} \Delta^{-1} \right) \Delta \\ & = C \left(p^{4/3} + r \ln \frac{4P(t)}{p} + r^{-1} \max \left\{ \frac{1}{t}, \frac{4C^* P(t)^{4/3}}{p} \right\} \right) \Delta; \end{aligned}$$

recall the definition of $\Delta = \Delta(t)$ in (1.22). We choose the parameters p and r in such a way that the terms in the sum on the right-hand side of this estimate are of the same order in $P(t)$,

$$p = P(t)^{4/11}, \quad r = P(t)^{16/33};$$

without loss of generality, $P(t) \geq 1$ so that $p \leq P(t)$, otherwise we replace $P(t)$ by $P(t) + 1$. Since P is nondecreasing and by Theorem 1.1, $\lim_{t \rightarrow T} P(t) = \infty$ if $T < \infty$, there exists a unique $T^* \in]0, T[$ such that $1/t \leq 4C^* P(t)^{4/3}/p = 4C^* P(t)^{32/33}$ for $t \geq T^*$. Hence for $t \geq T^*$,

$$|V(t) - V(t - \Delta)| \leq CP(t)^{16/33} \ln P(t) \Delta.$$

Thus, for any $\varepsilon > 0$ there exists a constant $C > 0$ such that

$$|V(t) - V(t - \Delta)| \leq CP(t)^{16/33+\varepsilon} \Delta, \quad t \geq T^*. \tag{1.29}$$

Let $t > T^*$ and define $t_0 := t$ and $t_{i+1} := t_i - \Delta(t_i)$ as long as $t_i \geq T^*$. Since

$$t_i - t_{i+1} = \Delta(t_i) \geq \Delta(t_0)$$

there exists $k \in \mathbb{N}$ such that

$$t_k < T^* \leq t_{k-1} < \cdots < t_0 = t.$$

Repeated application of (1.29) yields

$$\begin{aligned} |V(t) - V(t_k)| &\leq \sum_{i=1}^k |V(t_{i-1}) - V(t_i)| \\ &\leq C P(t)^{16/33+\varepsilon} \sum_{i=1}^k (t_{i-1} - t_i) \\ &\leq C P(t)^{16/33+\varepsilon} t. \end{aligned}$$

By the definition of P ,

$$P(t) \leq P(t_k) + C P(t)^{16/33+\varepsilon} t$$

so that for any $\delta > 0$ there exists a constant $C > 0$ such that

$$P(t) \leq C(1+t)^{33/17+\delta}, \quad t \in [0, T[,$$

and by Theorem 1.1 the proof is complete.

1.6.2. The Lions–Perthame proof. We present the ideas developed in [80] within the framework of classical solutions and use them to verify the continuation criterion in Proposition 1.10. Let

$$m_k(t, x) := m_k(f(t))(x), \quad M_k(t) := M_k(f(t)), \quad t \in [0, T[, x \in \mathbb{R}^3;$$

the right-hand terms were defined in Lemma 1.8. The field induced by the potential U is denoted by

$$F(t, x) := -\partial_x U(t, x).$$

The proof is split into a number of steps; constants denoted by C may depend on the initial datum, and their value may change from line to line. The order k of the moment to be bounded is specified below, but in any case, $k \geq 3$.

Step 1: A differential inequality for M_k . Using the Vlasov equation, integration by parts, Hölder's inequality, and Lemma 1.8 with $p = \infty$, $q = 1$, $k' = k - 1$, and hence

$r = (k + 3)/(k + 2)$ we obtain the following differential inequality

$$\begin{aligned}
 \left| \frac{d}{dt} M_k(t) \right| &= \left| \iint |v|^k (-v \cdot \partial_x f - F \cdot \partial_v f) dv dx \right| \\
 &= \left| k \iint |v|^{k-2} v \cdot F f dv dx \right| \leq k \iint |v|^{k-1} f dv |F| dx \\
 &\leq k \|F(t)\|_{k+3} \|m_{k-1}(t)\|_{(k+3)/(k+2)} \\
 &\leq C \|F(t)\|_{k+3} M_k(t)^{(k+2)/(k+3)}.
 \end{aligned} \tag{1.30}$$

Step 2: Straight forward estimates for the field. By Lemma P2(a) and Proposition 1.9 the estimate

$$\|F(t)\|_p \leq C, \quad t \in [0, T[, \tag{1.31}$$

holds for any $p \in]3/2, 15/4]$, and the constant C can be chosen to be independent of p . Hence the estimate

$$\tau^p \text{vol}\{x \in \mathbb{R}^3 \mid |F(t, x)| > \tau\} \leq C^p, \quad \tau > 0,$$

holds for all $3/2 < p \leq 15/4$, with C independent of p , so that the estimate also holds in the limiting case $p = 3/2$, which implies that

$$\|F(t)\|_{3/2, w} \leq C, \quad t \in [0, T[. \tag{1.32}$$

This estimate also follows from the fact that the mapping $L^1(\mathbb{R}^3) \ni \rho \mapsto \rho * 1/|\cdot|^2$ is of weak-type $(1, 3/2)$, cf. [108], Section V.1.2, Theorem 1.

Step 3: A representation formula for ρ and further estimates for the field. In order to proceed with the differential inequality (1.30) we need a suitable estimate for $\|F(t)\|_{k+3}$, which is not provided by Step 2. To this end we first derive a representation formula for the spatial density ρ . The Vlasov equation can be rewritten as follows:

$$\partial_t f + v \cdot \partial_x f = -\text{div}_v(fF).$$

We treat the right-hand side as an inhomogeneity and integrate this equation along the free streaming characteristics to obtain the following formula:

$$\begin{aligned}
 f(t, x, v) &= \mathring{f}(x - tv, v) - \int_0^t \text{div}_v(fF)(s, x + (s - t)v, v) ds \\
 &= \mathring{f}(x - tv, v) - \text{div}_v \int_0^t (fF)(s, x + (s - t)v, v) ds \\
 &\quad + \text{div}_x \int_0^t (s - t)(fF)(s, x + (s - t)v, v) ds.
 \end{aligned}$$

Integration with respect to v yields

$$\begin{aligned}\rho(t, x) &= \int \mathring{f}(x - tv, v) dv + \operatorname{div}_x \int_0^t (s - t) \int (fF)(s, x + (s - t)v, v) dv ds \\ &=: \rho_0(t, x) + \operatorname{div}_x \sigma(t, x).\end{aligned}$$

We split the field accordingly,

$$F = F_0 + F_1 := -\partial_x U_{\rho_0} - \partial_x U_{\operatorname{div} \sigma}.$$

The first term is easy to control. Because of the estimate

$$\rho_0(t, x) = \int \mathring{f}(x - tv, v) dv = t^{-3} \int \mathring{f}\left(X, \frac{x - X}{t}\right) dX \leq Ct^{-3},$$

the density contribution $\rho_0(t)$ is bounded on $[0, T[$ in any L^p -norm. Hence Lemma P2(a) implies that $\|F_0(t)\|_r$ is bounded on $[0, T[$ for any $r > 3/2$. To proceed with F_1 we need an auxiliary result, a consequence of the Calderon–Zygmund inequality.

LEMMA 1.12. *For any $p \in]1, \infty[$ there is a constant $c > 0$ such that for all $\sigma \in C_c^1(\mathbb{R}^3; \mathbb{R}^3)$,*

$$\|(\cdot/|\cdot|^3) * (\operatorname{div} \sigma)\|_p \leq c \|\sigma\|_p.$$

PROOF. Let $E := (\cdot/|\cdot|^3) * (\operatorname{div} \sigma)$. Integration by parts shows that for $i = 1, 2, 3$,

$$E_i(x) = \lim_{\varepsilon \rightarrow 0} \sum_{j=1}^3 (I_{1,\varepsilon}^{ij}(x) - I_{2,\varepsilon}^{ij}(x)),$$

where

$$\begin{aligned}I_{1,\varepsilon}^{ij}(x) &= \int_{|x-y|=\varepsilon} \frac{x_i - y_i}{|x - y|^3} \sigma_j(y) \frac{x_j - y_j}{|x - y|} d\Sigma(y), \\ I_{2,\varepsilon}^{ij}(x) &= \int_{|x-y|>\varepsilon} \partial_{y_j} \frac{x_i - y_i}{|x - y|^3} \sigma_j(y) dy.\end{aligned}$$

For $i \neq j$ the surface integral of the kernel in $I_{1,\varepsilon}^{ij}$ vanishes, and since $\sigma \in C_c^1(\mathbb{R}^3)$, $I_{1,\varepsilon}^{ij} \rightarrow 0$ for $\varepsilon \rightarrow 0$. For $i = j$ the integral of this kernel equals $4\pi/3$ so that $\sum_j I_{1,\varepsilon}^{ij} \rightarrow 4\pi\sigma/3$ for $\varepsilon \rightarrow 0$, uniformly on $\mathbb{R}^3$, and by the compact support assumption this convergence holds in L^p . The limit of $I_{2,\varepsilon}^{ij}$ can be estimated in the desired way by [108], Section II.4.2, Theorem 3, and the proof is complete. $\square$

Using this lemma for the term F_1 in the above splitting and the established bound on F_0 we arrive at the estimate

$$\|F(t)\|_{k+3} \leq C(1 + \|\sigma(t)\|_{k+3}), \quad t \in [0, T[. \quad (1.33)$$

In order to proceed we need another auxiliary result.

LEMMA 1.13. *For all functions $g \in L^1 \cap L^\infty(\mathbb{R}^3)$ and $h \in L_w^{3/2}(\mathbb{R}^3)$,*

$$\int |gh| \, dx \leq 3 \left(\frac{3}{2}\right)^{2/3} \|g\|_1^{1/3} \|g\|_\infty^{2/3} \|h\|_{3/2, w}.$$

PROOF. For any $\tau > 0$ the “layer cake representation” [74], Section 1.13, implies that

$$\int_{|h|>\tau} |h| \, dx \leq 3 \|h\|_{3/2, w}^{3/2} \tau^{-1/2},$$

and hence

$$\begin{aligned} \int |gh| \, dx &= \int_{|h| \leq \tau} |gh| \, dx + \int_{|h| > \tau} |gh| \, dx \\ &\leq \tau \|g\|_1 + 3 \|h\|_{3/2, w}^{3/2} \tau^{-1/2} \|g\|_\infty. \end{aligned}$$

If we choose $\tau := \|h\|_{3/2, w} (3/2)^{2/3} (\|g\|_\infty / \|g\|_1)^{2/3}$ the assertion follows. $\square$

Step 4: Gronwall estimate for M_k . In order to derive a Gronwall inequality for the moment M_k we need to estimate $\|\sigma(t)\|_{k+3}$ in terms of a moment, cf. Steps 1 and 3. We fix some time $t_0 \in]0, T[$, to be chosen in a suitable way later on. Then for any $t \in]t_0, T[$ we have by the definition of the quantity σ in Step 2,

$$\begin{aligned} \|\sigma(t)\|_{k+3} &= \left\| \int_0^t (s-t) \int (fF)(s, \cdot + (s-t)v, v) \, dv \, ds \right\|_{k+3} \\ &\leq \left\| \int_0^{t_0} \dots \right\|_{k+3} + \left\| \int_{t_0}^t \dots \right\|_{k+3} =: I_1 + I_2. \end{aligned}$$

By Lemma 1.13, equation (1.32), a change of variables, and the boundedness of f ,

$$\int (|F|f)(s, x + (s-t)v, v) \, dv \leq C(t-s)^{-2} \left(\int f(s, x + (s-t)v, v) \, dv \right)^{1/3}.$$

We use this estimate and Lemma 1.8 for the term I_1 ,

$$\begin{aligned} I_1 &\leq C \left\| \int_0^{t_0} (t-s)^{-1} \left(\int f(s, \cdot + (s-t)v, v) dv \right)^{1/3} ds \right\|_{k+3} \\ &\leq C \int_0^{t_0} (t-s)^{-1} \left\| \int f(s, \cdot + (s-t)v, v) dv \right\|_{(k+3)/3}^{1/3} ds \\ &\leq C \int_0^{t_0} (t-s)^{-1} M_k(s)^{1/(k+3)} ds. \end{aligned}$$

This estimate is good as long as we stay away from the singularity of the integrand, but for s close to t , i.e., on the interval $[t_0, t]$ we have to argue differently. We fix some parameter $3/2 < r \leq 15/4$, to be specified later, and its dual exponent defined by $1/r + 1/r' = 1$. Then by Hölder's inequality and (1.31),

$$\begin{aligned} I_2 &\leq \left\| \int_{t_0}^t (t-s) \left(\int |F(s, \cdot + (s-t)v)|^r dv \right)^{1/r} \right. \\ &\quad \times \left. \| \hat{f} \|_{\infty}^{(r'-1)/r'} \left(\int f(s, \cdot + (s-t)v, v) dv \right)^{1/r'} ds \right\|_{k+3} \\ &\leq C \int_{t_0}^t (t-s)^{1-3/r} \left\| \int f(s, \cdot + (s-t)v, v) dv \right\|_{(k+3)/r'}^{1/r'} ds. \end{aligned}$$

Let $l > 0$ be such that $(l+3)/3 = (k+3)/r'$; such a choice is possible since

$$\frac{k+3}{r'} \geq \frac{6}{r'} > 1.$$

Applying Lemma 1.8 to the v -integral in the last estimate we conclude that

$$I_2 \leq C \int_{t_0}^t (t-s)^{1-3/r} M_l(s)^{1/(k+3)} ds.$$

In order to continue it is convenient to have M_k nondecreasing in t so we replace $M_k(t)$ by $\sup_{0 \leq s \leq t} M_k(s)$. Collecting the estimates for I_1 and I_2 we obtain by (1.33) the following estimate for the field:

$$\begin{aligned} &\|F(t)\|_{k+3} \\ &\leq C \left(1 + M_k(t)^{1/(k+3)} \ln \frac{t}{t-t_0} + M_l(t)^{1/(k+3)} (t-t_0)^{2-3/r} \right). \end{aligned} \quad (1.34)$$

If on the other hand, we integrate the differential inequality from Step 1 we obtain the estimate

$$M_l(s) \leq M_l(0) + C \sup_{0 \leq s \leq t} \|F(s)\|_{l+3} \int_0^s M_l(\tau)^{(l+2)/(l+3)} d\tau,$$

which implies that for $0 \leq s \leq t$,

$$M_l(s) \leq C \left(1 + t^{l+3} \sup_{0 \leq s \leq t} \|F(s)\|_{l+3}^{l+3} \right). \quad (1.35)$$

In order to close the Gronwall loop for the quantity M_k we must estimate the L^{l+3} -norm of the field $F(s)$ in terms of $M_k(s)$, which by Lemma 1.8 means that we must estimate it in terms of the $L^{(k+3)/3}$ -norm of the spatial density. The way to do the latter is to again use Lemma P2(a), so we now must adjust the exponents k, l, r properly. We need that

$$\frac{3}{k+3} + \frac{2}{3} = 1 + \frac{1}{l+3}, \quad \text{i.e., } l+3 = \left(\frac{3}{k+3} - \frac{1}{3} \right)^{-1} = \frac{3(k+3)}{6-k}.$$

Since l must be positive, k must satisfy the restriction

$$3 \leq k < 6.$$

On the other hand, we have to observe the relation between k, l, r' used above

$$\frac{k+3}{6-k} = \frac{l+3}{3} = \frac{k+3}{r'}$$

which implies that

$$r' = 6 - k, \quad \text{i.e., } \frac{1}{r} = 1 - \frac{1}{r'} = \frac{5-k}{6-k}.$$

Now we recall that for the estimates above the restriction

$$\frac{3}{2} < r \leq \frac{15}{4}, \quad \text{i.e., } \frac{3}{2} < \frac{6-k}{5-k} \leq \frac{15}{4}$$

was required. We end up with the result that the exponents k, l, r can be chosen such that all the relations introduced so far do indeed hold iff

$$3 < k \leq \frac{51}{11}.$$

This is the range of exponents for which we now establish a bound on M_k . By Lemma P2(a) and Lemma 1.8,

$$\|F(s)\|_{l+3} \leq C \|\rho(s)\|_{(k+3)/3} \leq C M_k(t)^{3/(k+3)}, \quad 0 \leq s \leq t < T,$$

and by (1.34) and (1.35),

$$\begin{aligned} & \|F(t)\|_{k+3} \\ & \leq C \left(1 + M_k(t)^{1/(k+3)} \ln \frac{t}{t-t_0} + M_k(t)^{3(l+3)/(k+3)^2} t^{l+3} (t-t_0)^{2-3/r} \right). \end{aligned}$$

We have to examine the various exponents. Clearly,

$$2 - \frac{3}{r} > 0.$$

Since $1/r' > 1/3$ we have $l+3 > k+3$ and hence

$$\frac{1}{k+3} - \frac{3(l+3)}{(k+3)^2} = \frac{1}{k+3} \left(1 - \frac{3(l+3)}{k+3} \right) < 0.$$

By monotonicity there exists a unique time $t^* \in]0, T[$ such that

$$M_k(t)^{1/(k+3)-3(l+3)/(k+3)^2} < t^{2-3/r}, \quad t \geq t^*;$$

without loss of generality $M_k(0) > 0$. For $t \geq t^*$ we choose $t_0 \in]0, t[$ such that

$$(t-t_0)^{2-3/r} = M_k(t)^{1/(k+3)-3(l+3)/(k+3)^2},$$

and hence

$$\|F(t)\|_{k+3} \leq C t^{l+3} \ln t M_k(t)^{1/(k+3)} \ln M_k(t).$$

If we insert this estimate into the integrated differential inequality from Step 1 we finally arrive at the estimate

$$M_k(t) \leq C + C \int_{t^*}^t s^{l+3} \ln s M_k(s) \ln M_k(s) ds, \quad t \in [t^*, T].$$

Hence M_k is bounded on bounded time intervals, and by Proposition 1.10 the proof is complete.

Concluding remarks. (a) In addition to being more elementary the Pfaffelmoser and Schaeffer proof yields better bounds on $P(t)$ and $\|\rho(t)\|_\infty$. Using a somewhat different refinement of the original Pfaffelmoser proof Horst [60] showed that for any $\delta > 0$,

$$P(t) \leq C(1+t)^{1+\delta}, \quad \|\rho(t)\| \leq C(1+t)^{3+\delta}, \quad t \geq 0.$$

(b) The Pfaffelmoser and Schaeffer proof has been employed for the Vlasov–Poisson system in a spatially periodic, plasma physics setting [12] and in a cosmological setting [101].

(c) The Lions and Perthame ideas have the greater potential to generalize to related systems. In [14] global existence was established for the Vlasov–Fokker–Planck–Poisson system. In this system collisional effects are included by a linear approximation of the Boltzmann collision operator, and instead of the Vlasov equation (1.1) the so-called Vlasov–Fokker–Planck equation

$$\partial_t f + v \cdot \partial_x f - (\partial_x U + \beta v) \cdot \partial_v f = 3\beta f + \sigma \Delta_v f$$

is coupled to (1.2), (1.3), where $\beta, \sigma > 0$. Notice that the method of characteristics does not apply to this equation. The Lions and Perthame techniques were also successful for a version of the Vlasov–Poisson system which includes a damping term modeling the fact that charges in motion radiate energy [68].

(d) A proof based on the ideas of Lions and Perthame but using moments with respect to x and v is given in [27].

(e) The above global existence results extend easily to the plasma physics case with several particle species, with a fixed ion background, or with a fixed exterior field. If, on the other hand, the system is considered on a spatial domain with boundary, where a variety of boundary conditions for the particles can be posed like specular reflexion, absorption, or an inflow boundary condition, then the situation changes drastically, and in general not even a local existence and uniqueness result is known. We refer to [38–40, 62, 111] for results on Vlasov-type systems with boundary conditions.

(f) For the plasma physics case of the relativistic Vlasov–Poisson system and for the relativistic Vlasov–Maxwell system no global classical existence result for general data has been proven yet, cf. also the concluding remarks of Section 1.5. In addition to the papers mentioned in Section 1.1 we mention [86] where Pallard introduced significant new ideas for the Vlasov–Maxwell system.

(g) The ideas of Pallard have very recently been exploited by Calogero [17] to prove global existence of classical solutions for general data to the Vlasov–Nordström system

$$\begin{aligned} \partial_t f + \frac{v}{\sqrt{1+|v|^2}} \cdot \partial_x f - \left[\left(\partial_t \phi + \frac{v}{\sqrt{1+|v|^2}} \cdot \partial_x \phi \right) v + \frac{1}{\sqrt{1+|v|^2}} \partial_x \phi \right] \cdot \partial_v f \\ = 0, \\ \partial_t^2 \phi - \Delta \phi = -e^{4\phi} \int f \frac{dv}{\sqrt{1+|v|^2}}. \end{aligned}$$

In view of the notoriously difficult Vlasov–Einstein system the Vlasov–Nordström system can serve as a toy model to gain experience with relativistic, gravitationally interacting particle ensembles. It is much simpler than the former, physically correct system, but the Vlasov equation is relativistic in the sense that its characteristic system are the geodesic equations in the metric $e^{2\phi} \text{diag}(-1, 1, 1, 1)$, and the field equation for the function $\phi = \phi(t, x)$ is hyperbolic. Notice that compared to the Vlasov–Maxwell system the source term in the field equation here is of lower order in v , which is important for the success of the proof.

1.7. Asymptotic behavior

Due to Proposition 1.9 the global existence proofs go through and give the same bounds for both the plasma physics and the gravitational case. The latter fact is not satisfactory, since in the plasma physics case the particles repulse each other – in the case of several particle species with charges of different sign this is true at least on the average – and thus the spatial density should decay as $t \rightarrow \infty$. In the present subsection we present results of this type, which also form a link to the next chapter where we study the stability of steady states. The main tool are certain identities satisfied by a solution; the second one was introduced and exploited in [63,87] and put into a larger context in [23].

THEOREM 1.14. *Let f be a classical solution of the Vlasov–Poisson system with nonnegative initial datum $\hat{f} \in C_c^1(\mathbb{R}^6)$. Then the following identities hold for all times:*

$$\frac{1}{2} \frac{d^2}{dt^2} \int \int |x|^2 f(t, x, v) dv dx = 2E_{\text{kin}}(f(t)) + E_{\text{pot}}(f(t)), \quad (1.36)$$

$$\frac{d}{dt} \left[\frac{1}{2} \int \int |x - tv|^2 f(t, x, v) dv dx + t^2 E_{\text{pot}}(f(t)) \right] = t E_{\text{pot}}(f(t)). \quad (1.37)$$

PROOF. We start by proving yet another identity which is sometimes referred to as the dilation identity, cf. [35]. By the Vlasov equation,

$$\begin{aligned} \frac{d}{dt} \int \int x \cdot v f(t, x, v) dv dx &= \int \int x \cdot v (-v \cdot \partial_x f + \partial_x U \cdot \partial_v f) dv dx \\ &= \int \int |v|^2 f dv dx - \int x \cdot \partial_x U \rho dx. \end{aligned}$$

Now the formulas for the potential and its gradient imply that

$$\begin{aligned} \int x \cdot \partial_x U \rho dx &= \gamma \int \int x \frac{x - y}{|x - y|^3} \rho(t, y) \rho(t, x) dy dx \\ &= \frac{1}{2} \gamma \int \int (x - y) \frac{x - y}{|x - y|^3} \rho(t, y) \rho(t, x) dy dx \\ &= -\frac{1}{2} \int U \rho dx = -E_{\text{pot}}(f(t)). \end{aligned}$$

Hence

$$\frac{d}{dt} \int \int x \cdot v f(t, x, v) dv dx = 2E_{\text{kin}}(f(t)) + E_{\text{pot}}(f(t)). \quad (1.38)$$

Together with

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \int \int |x|^2 f \, dv \, dx &= \frac{1}{2} \int \int |x|^2 (-v \cdot \partial_x f + \partial_x U \cdot \partial_v f) \, dv \, dx \\ &= \int \int x \cdot v f \, dv \, dx, \end{aligned}$$

this implies the identity (1.36). Since the total energy $\mathcal{H} := E_{\text{kin}} + E_{\text{pot}}$ is conserved,

$$\begin{aligned} &\frac{d}{dt} \left[\frac{1}{2} \int \int |x - tv|^2 f(t, x, v) \, dv \, dx + t^2 E_{\text{pot}}(f(t)) \right] \\ &= \frac{d}{dt} \left[\frac{1}{2} \int \int |x|^2 f \, dv \, dx + t^2 \mathcal{H}(f(t)) - \frac{1}{2} t \frac{d}{dt} \int \int |x|^2 f \, dv \, dx \right] \\ &= 2t \mathcal{H}(f(t)) - \frac{1}{2} t \frac{d^2}{dt^2} \int \int |x|^2 f \, dv \, dx = t E_{\text{pot}}(f(t)), \end{aligned}$$

and the proof is complete. $\square$

In the plasma physics case the identity (1.37) implies that solutions decay.

COROLLARY 1.15. *In the plasma physical case $\gamma = -1$ there exists for every solution of the Vlasov–Poisson system with initial datum as above a constant $C > 0$ such that the following estimates hold for all $t \geq 0$:*

$$\|\partial_x U(t)\|_2 \leq C(1+t)^{-1/2}, \quad (1.39)$$

$$\int \int |v - x/t|^2 f(t, x, v) \, dv \, dx \leq C(1+t)^{-1}, \quad (1.40)$$

$$\|\rho(t)\|_{5/3} \leq C(1+t)^{-3/5}. \quad (1.41)$$

PROOF. Since $\gamma = -1$, the quantity

$$g(t) := t^2 E_{\text{pot}}(f(t)) \geq 0$$

is nonnegative. The identity (1.37) takes the form

$$\frac{d}{dt} \left[\frac{1}{2} \int \int |x - tv|^2 f(t, x, v) \, dv \, dx + g(t) \right] = \frac{g(t)}{t}, \quad t > 0.$$

Integration of this identity from 1 to $t \geq 1$ yields

$$\frac{1}{2} \int \int |x - tv|^2 f(t, x, v) \, dv \, dx + g(t) = C + \int_1^t \frac{g(s)}{s} \, ds \quad (1.42)$$

for some constant $C > 0$ which depends on $f(1)$. We drop the double integral and apply Gronwall's lemma to the resulting inequality to obtain the estimate $g(t) \leq Ct$ for $t \geq 1$, and this proves (1.39). Insertion of the estimate for g into (1.42) proves (1.40). To obtain the estimate (1.41) we repeat the argument from the proof of Lemma 1.8, but instead of splitting the ρ integral according to $|v| < / > R$ we split according to $|v - x/t| < / > R$, and instead of the kinetic energy density $\int |v|^2 f \, dv$ we use the quantity $\int |v - x/t|^2 f \, dv$. $\square$

It should be emphasized that this type of decay remains true if the plasma consists of several species of particles with charges of different sign. In particular, this shows that in the plasma physics case the Vlasov–Poisson system as stated above does not have stationary solutions.

In the stellar dynamics case, Theorem 1.14 yields a dispersion result for solutions with positive energy. This was first observed in [24].

COROLLARY 1.16. *Consider a solution f of the Vlasov–Poisson system in the stellar dynamics case $\gamma = 1$ with positive energy: $\mathcal{H}(f) = E_{\text{kin}}(f) + E_{\text{pot}}(f) > 0$. Then there exist constants $C_1, C_2 > 0$ which depend on $\mathcal{H}(f)$, $\|f\|_1$, $\|f\|_\infty$ such that for all sufficiently large times,*

$$C_1 t^2 \leq \iint |x|^2 f(t, x, v) \, dv \, dx \leq C_2 t^2.$$

In particular,

$$\sup\{|x| \mid (x, v) \in \text{supp } f(t)\} \geq \left(\frac{C_1}{\|f\|_1} \right)^{1/2} t.$$

PROOF. By Proposition 1.9 and since the potential energy is negative,

$$0 < 2\mathcal{H}(f) \leq 2E_{\text{kin}}(f(t)) + E_{\text{pot}}(f(t)) \leq C_2,$$

and the assertion follows from the identity (1.36). $\square$

The corollary implies in particular that any stationary solution in the stellar dynamics case must have negative energy.

As pointed out for example in the remarks following Proposition 1.10, global existence does not hold for the relativistic Vlasov–Poisson system (1.4), (1.2), (1.3) in the gravitational case [29]. This result can be seen using an identity similar to the above.

THEOREM 1.17. *Let f be a spherically symmetric, classical solution to the gravitational, relativistic Vlasov–Poisson system (1.4), (1.2), (1.3) with nonnegative initial datum $f \in C_c^1(\mathbb{R}^6)$ and with negative energy*

$$\mathcal{H}(t) := \frac{1}{2} \iint \sqrt{1 + |v|^2} f(t, x, v) \, dv \, dx - \frac{1}{8\pi} \int |\partial_x U(t, x)|^2 \, dx = \mathcal{H}(0) < 0.$$

Then this solution blows up in finite time.

PROOF. A computation analogous to the one leading to equation (1.38) yields the following relativistic dilation identity, which holds for any classical solution as long as it exists,

$$\frac{d}{dt} \iint x \cdot v f(t, x, v) dv dx = \mathcal{H}(t) - \iint \frac{1}{\sqrt{1+|v|^2}} f(t, x, v) dv dx.$$

This implies that, for $t \geq 0$,

$$\iint x \cdot v f(t, x, v) dv dx \leq C + t\mathcal{H}(0).$$

Moreover, with

$$j(t, x) := \int \frac{v}{\sqrt{1+|v|^2}} f(t, x, v) dv$$

and the previous estimate, we find that

$$\begin{aligned} \frac{d}{dt} \iint |x|^2 \sqrt{1+|v|^2} f(t) dv dx \\ = 2 \iint x \cdot v f(t) dv dx - \int |x|^2 (\partial_x U \cdot j)(t, x) dx \\ \leq C + 2\mathcal{H}(0)t - \int |x|^2 (\partial_x U \cdot j)(t, x) dx. \end{aligned}$$

Due to spherical symmetry, using (1.10),

$$\left| \int |x|^2 (\partial_x U \cdot j)(t, x) dx \right| \leq \|f(t)\|_1 \int |j(t, x)| dx \leq \|f(t)\|_1^2 = C$$

and hence

$$0 \leq \iint |x|^2 \sqrt{1+|v|^2} f(t) dv dx \leq C(1+t) + \mathcal{H}(0)t^2.$$

But this estimate cannot hold for all $t > 0$, since by assumption, $\mathcal{H}(0) < 0$. □

Concluding remarks. (a) In [93] the decay estimates from Corollary 1.15 for the plasma physics case are used as input in the Pfaffelmoser and Schaeffer proof to obtain the improved estimates

$$P(t) \leq C(1+t)^{2/3}, \quad \|\rho(t)\|_\infty \leq C(1+t)^2, \quad t \geq 0.$$

(b) In the plasma physics case, at least for the case of only one particle species, one may *conjecture* that the decay estimate

$$\|\rho(t)\|_{\infty} \leq Ct^{-3}, \quad t \geq 0,$$

holds. If true, this decay rate would be sharp,

$$t^3 \|\rho(t)\|_{\infty} \not\rightarrow 0, \quad t \rightarrow \infty,$$

since with such a decay the velocities remain bounded, the diameter of the spatial support grows at most linearly in t , but the mass is conserved. For small data the above decay rate does hold. Using the techniques from Section 1.4 one can show the following: If a solution satisfies the decay estimate $\|\rho(t)\|_{\infty} \leq Ct^{-\alpha}$ with some $\alpha > 2$ then any solution starting in a small neighborhood satisfies the decay estimate with $\alpha = 3$. For spherically symmetric solutions the estimate $\|\rho(t)\|_{\infty} \leq Ct^{-3} \ln(1+t)$ was established in [59]. In space dimension one, $\|\rho(t)\|_{\infty} \leq Ct^{-1}$, cf. [10].

(c) Since the analogue of Theorem 1.1 holds for the relativistic Vlasov–Poisson system, blow-up means that ρ blows up in the L^{∞} -norm. Moreover, it is easy to see that in the spherically symmetric situation this blow-up has to occur at the origin. It is an interesting open problem to show that this blow-up behavior persists without the symmetry assumption. It is maybe not of physical but of mathematical interest that such blow-up results hold for the (nonrelativistic) Vlasov–Poisson system in space dimensions greater than or equal to 4, cf. [58,72].

2. Stability

2.1. Introduction – steady states, stability and energy-Casimir functionals

The question of which steady states of the Vlasov–Poisson system are stable in the gravitational case has over decades received a lot of attention in the astrophysics literature, and it still is an active field of research in astrophysics. The stability problem is of course also of considerable importance in plasma physics, and we will make occasional remarks on this case, but except for such remarks we consider only the gravitational case $\gamma = 1$ in this section. The corresponding results in the plasma physics case are in comparison easy to obtain. The results of this chapter originate in the collaboration of Y. Guo and the author [41,42,46–49,94,96–98]. They are presented here in a unified way.

Before entering into a discussion of the stability question there arises a presumably simpler question: Does the system *have* steady states?

2.1.1. A strategy to construct steady states. If $U_0 = U_0(x)$ is a time independent potential then the local or particle energy

$$E = E(x, v) := \frac{1}{2}|v|^2 + U_0(x) \tag{2.1}$$

is constant along solutions of the characteristic system

$$\dot{x} = v, \quad \dot{v} = -\nabla U_0(x).$$

Hence E as well as any function of E solves the Vlasov equation for the potential U_0 . This leads to the following ansatz for a stationary solution:

$$f_0(x, v) = \phi(E) = \phi(E(x, v)),$$

where ϕ is a suitably chosen function. By this ansatz the Vlasov equation is satisfied. The spatial density ρ_0 becomes a functional of the potential U_0 , and in order to obtain a self-consistent stationary solution of the Vlasov–Poisson system the remaining, semilinear Poisson equation must be solved, cf. (2.2). If a solution exists then the above ansatz *defines* a steady state with induced potential U_0 . However, not just any solution obtained in this manner is acceptable. The resulting phase space density f_0 needs to have finite mass and possibly finite support in space; it should be noticed that the semilinear Poisson equation has to be solved on the whole space $\mathbb{R}^3$, since it is a priori not known where the support of the steady state will end and whether it will be bounded in the first place. It can be shown that these properties can hold only if the distribution vanishes for large values of the local energy E , cf. [102], Theorem 2.1. It turns out to be convenient to slightly reformulate the problem.

Steady state existence problem. Specify conditions on a measurable function $\phi : \mathbb{R} \rightarrow [0, \infty[$ with $\phi(\eta) = 0$ for $\eta < 0$ such that there exists a cut-off energy $E_0 \in \mathbb{R}$ and a solution U_0 of the semilinear elliptic problem

$$\Delta U_0 = 4\pi \int \phi\left(E_0 - \frac{1}{2}|v|^2 - U_0\right) dv, \quad \lim_{|x| \rightarrow \infty} U_0(x) = 0, \quad (2.2)$$

with

$$U_0(x) \geq E_0 \quad \text{for } |x| \text{ sufficiently large.}$$

If U_0 is a solution of this problem then up to regularity issues

$$f_0(x, v) := \phi\left(E_0 - \frac{1}{2}|v|^2 - U_0(x)\right)$$

defines a steady state which is compactly supported and hence, if for example f_0 is bounded, has finite mass.

We do not enter more deeply into the matter of the existence of steady states for two reasons: Firstly, in our stability analysis we actually prove the existence of stable steady states. Secondly, steady states of the form discussed above must a posteriori be spherically symmetric so that the semilinear Poisson equation (2.2) becomes an ordinary differential equation with respect to the radial variable $r = |x|$, and its analysis does not really fit into

the present treatise. In order to demonstrate the wide variety of possible steady states we present some of the known results without proofs. If U_0 is spherically symmetric the square of the modulus of angular momentum

$$L := |x \times v|^2 \quad (2.3)$$

is conserved along characteristics, and the distribution function can be taken to depend on E and L . The so-called polytropic ansatz

$$f_0(x, v) = (E_0 - E)_+^k L^l$$

with $k > -1$, $l > -1$, $k + l + 1/2 > 0$, $k < 3l + 7/2$, leads to steady states with finite mass and compact support, cf. [9]. In the limiting case $k = 3l + 7/2$ the mass is still finite but the support is the whole space, and for $k > 3l + 7/2$ the resulting steady state has infinite mass.

The dependence on L entails additional problems in the stability analysis which we will comment on later. In the sequel we restrict ourselves to steady states which depend only on the particle energy, so-called isotropic states. In [102], Theorem 3.1, it is shown that the approach above leads to a steady state with finite mass and compact support, provided $\phi \in L_{\text{loc}}^\infty(\mathbb{R})$ and

$$\phi(E) = c(E_0 - E)^k + O((E_0 - E)^{k+\delta}) \quad \text{as } E \rightarrow E_0 -$$

with parameters $0 < k < 3/2$ and $\delta, c > 0$. This result covers the isotropic polytropes with $0 < k < 3/2$. More recently, a similar generalization of the isotropic polytropes with $0 < k < 7/2$ has been found, cf. [55], but then a condition on the global behavior of ϕ is required. We will repeatedly encounter the threshold $k = 3/2$ in what follows.

The steady states mentioned so far are spherically symmetric, and much less is known about the existence of steady states with less symmetry. In addition to the results in [49, 94], which will be discussed in the stability context, we mention that axially symmetric steady states can be obtained as perturbations of spherically symmetric ones via the implicit function theorem [95].

For the plasma physics case the system as stated above has no steady states, cf. Section 1, Corollary 1.15. In order to have steady states in the plasma physics case one needs to include an exterior field or a fixed ion background or to consider the system on a bounded domain with appropriate boundary conditions. In these situations steady states are fairly easy to obtain, cf. [11, 91]. The problem becomes more challenging if one is interested in steady states with a nontrivial magnetic field [45].

2.1.2. Stability via linearization? We do not wish to enter a general discussion of possible stability concepts, for which we refer to [56]. An often successful strategy to analyze the stability properties of some steady state of a dynamical system is linearization. For the Vlasov–Poisson system this approach is often followed in the astrophysics literature. We briefly review some of the arguments which can for example be found in the monographs

[13,26], where the interested reader will find many further references. Assume that f_0 is a steady state with induced spatial density ρ_0 and potential U_0 , and let

$$f = f_0 + g, \quad \rho = \rho_0 + \sigma, \quad U = U_0 + W,$$

denote a solution of the time dependent problem which starts close to the steady state. We obtain the linearized system (linearized about f_0) if we substitute the above into the Vlasov–Poisson system, use the fact that (f_0, ρ_0, U_0) satisfies the system, and drop the quadratic term in the Vlasov equation

$$\partial_t g + v \cdot \partial_x g - \nabla U_0 \cdot \partial_v g = \partial_v f_0 \cdot \partial_x W,$$

$$\Delta W = 4\pi\sigma, \quad \lim_{|x| \rightarrow \infty} W(t, x) = 0,$$

$$\sigma(t, x) = \int g(t, x, v) dv.$$

Now we assume that the steady state is of the form discussed above, i.e., $f_0(x, v) = \phi(E)$. If we use the abbreviation

$$D := v \cdot \partial_x - \nabla U_0 \cdot \partial_v,$$

observe that $DE = 0$, and substitute the formula for the Newtonian potential we obtain the equivalent equation

$$\partial_t g + D \left[g + \phi'(E) \iint \frac{g(t, y, w)}{|x - y|} dw dy \right] = 0.$$

What we have in mind here is only an exploratory calculation so we assume that everything is as regular as necessary for our manipulations. With some abuse of notation we make the ansatz

$$g(t, x, v) = e^{\lambda t} g(x, v)$$

so that the linearized problem takes the form

$$\lambda g + D \left[g + \phi'(E) \iint \frac{g(y, w)}{|x - y|} dw dy \right] = 0. \quad (2.4)$$

If $\operatorname{Re} \lambda < 0$ for all solutions (λ, g) then the steady state f_0 is expected to be stable, if $\operatorname{Re} \lambda > 0$ for one solution then it should be unstable. For finite-dimensional dynamical systems (ordinary differential equations) this expectation is of course justified by rigorous theorems. However, for infinite-dimensional dynamical systems such as the Vlasov–Poisson system no general such results exist.

There is however a more specific problem with an attempt to prove stability via the above spectral analysis of the linearized system. We split g into its even and odd parts in v , i.e.,

$$g = g_+ + g_-, \quad \text{where } g_{\pm}(x, v) := \frac{1}{2}[g(x, v) \pm g(x, -v)].$$

If we substitute this into the eigenvalue equation (2.4) and group together the even and odd parts we see that we have to solve the system of equations

$$\begin{aligned} \lambda g_+ + Dg_- &= 0, \\ \lambda g_- + Dg_+ + D\left[\phi'(E) \iint \frac{g_+(y, w)}{|x - y|} dw dy\right] &= 0. \end{aligned}$$

We eliminate g_+ , and it remains to investigate the equation

$$\lambda^2 g_- = D^2 g_- + D\left[\phi'(E) \iint \frac{Dg_-(y, w)}{|x - y|} dw dy\right] = 0,$$

where only solutions $g_- = g_-(x, v)$ are relevant which are odd in v . But if the pair (λ, g_-) solves this equation then so does the pair $(-\lambda, g_-)$. Hence as far as stability is concerned the best we may hope for is that all the eigenvalues λ are purely imaginary. Since in this situation one can in general draw no conclusion about the nonlinear stability of the steady state, not even for finite-dimensional dynamical systems, we do not pursue linearization any further.

For the plasma physics case a linearized analysis based on conserved quantities instead of spectral properties is carried out in [11].

2.1.3. Energy-Casimir functionals. As noted in Section 1.5, the Vlasov–Poisson system conserves energy: The functional

$$\mathcal{H}(f) := E_{\text{kin}}(f) + E_{\text{pot}}(f) = \frac{1}{2} \int |v|^2 f(x, v) dv dx - \frac{1}{8\pi} \int |\nabla U_f(x)|^2 dx$$

is constant along solutions. A natural approach to the stability question for a conservative system is to use the energy as a Lyapunov function. This idea meets an immediate obstacle: For the Lyapunov approach to work the steady state must first of all be a critical point of the energy, but in the present case the energy does not have critical points, i.e., the linear part in an expansion about any state f_0 with potential U_0 does not vanish,

$$\begin{aligned} \mathcal{H}(f) &= \mathcal{H}(f_0) + \iint \left(\frac{1}{2}|v|^2 + U_0\right)(f - f_0) dv dx \\ &\quad - \frac{1}{8\pi} \int |\nabla U_f - \nabla U_0|^2 dx. \end{aligned}$$

However, the characteristic flow corresponding to the Vlasov equation preserves phase space volume, cf. Section 1, Lemma 1.2(b), and hence for any reasonable function Φ the so-called *Casimir functional*

$$\mathcal{C}(f) := \iint \Phi(f(x, v)) \, dv \, dx$$

is conserved as well. If the energy-Casimir functional

$$\mathcal{H}_C := \mathcal{H} + \mathcal{C}$$

is expanded about an isotropic steady state

$$f_0(x, v) = \phi(E)$$

with the particle energy E defined as in (2.1), then

$$\begin{aligned} \mathcal{H}_C(f) = \mathcal{H}_C(f_0) &+ \iint (E + \Phi'(f_0))(f - f_0) \, dv \, dx \\ &- \frac{1}{8\pi} \int |\nabla U_f - \nabla U_0|^2 \, dx + \frac{1}{2} \iint \Phi''(f_0)(f - f_0)^2 \, dv \, dx + \dots \end{aligned} \quad (2.5)$$

At least formally, we can choose Φ such that f_0 is a critical point of $\mathcal{H}_C$, namely $\Phi' = -\phi^{-1}$, provided ϕ is invertible. In more abstract terms we can say that the Hamiltonian $\mathcal{H}$ does not have critical points when we take as state space the space of all phase space densities f , but given a Casimir functional defined as above the corresponding steady state is a critical point of the Hamiltonian restricted to the manifold which is defined by the constraint $\mathcal{C}(f) = \mathcal{C}(f_0)$. A mostly formal discussion of this energy-Casimir approach in the context of so-called degenerate Hamiltonian or Lie–Poisson systems can be found in [56]. We make no use of this abstract background of our problem.

The question now is whether the quadratic term in the expansion (2.5) is positive (or negative) definite. As was noted in Section 2.1.1, in order for the steady state to have finite total mass the function ϕ must vanish above a certain cut-off energy. For ϕ^{-1} to exist ϕ should thus be decreasing, at least on its support. But then Φ'' is positive and the quadratic part in the expansion indefinite. Since one would like to use this quadratic part for defining the concept of distance or neighborhood, the method seems to fail.

If the issue is the stability of a plasma, the sign in front of the potential energy difference in the expansion (2.5) is reversed, and up to some technicalities stability follows, cf. [92]. The technical difficulties are among other things due to the fact that ϕ can at best be invertible on its support which is bounded from above by the cut-off energy.

2.1.4. A variational problem and stability. As noted above, certain steady states of the Vlasov–Poisson system are critical points of an energy-Casimir functional. At the same time the quadratic term in the Taylor expansion of this functional about such a steady state looks at first glance indefinite, which bodes ill for a stability analysis.

In this section we reverse our strategy in the following sense: We do not start with a given steady state whose stability we want to investigate, but instead we start with an energy-Casimir functional, i.e., with a function Φ defining the Casimir part, and we ask whether this functional attains its minimum on a suitable set of states f . Such a minimizer, if it exists, is a critical point of the energy-Casimir functional and hence should be a steady state, and its minimizing property can hopefully lead to a stability assertion.

Hence let $\Phi : [0, \infty[\rightarrow [0, \infty[$ be given – the necessary assumptions on this function are stated below. We investigate two closely related, but different variational problems, both of which have their merits. The difference between the two problems lies in the role of the Casimir functional – in the first formulation it is part of the functional to be minimized, in the second one it is part of the following constraint.

VARIATIONAL PROBLEM – VERSION 1. *Minimize the energy-Casimir functional $\mathcal{H}_C = \mathcal{H} + \mathcal{C}$ under a mass constraint, i.e., prove that the functional $\mathcal{H}_C$ has a minimizer $f_0 \in \mathcal{F}_M$,*

$$\mathcal{H}_C(f) \geq \mathcal{H}_C(f_0) \quad \text{for all } f \in \mathcal{F}_M,$$

where the constraint set is defined as

$$\mathcal{F}_M := \left\{ f \in L^1_+(\mathbb{R}^6) \mid \iint f \, dv \, dx = M, E_{\text{kin}}(f) + \mathcal{C}(f) < \infty \right\}.$$

VARIATIONAL PROBLEM – VERSION 2. *Minimize the energy functional $\mathcal{H}$ under a mass-Casimir constraint, i.e., prove that the functional $\mathcal{H}$ has a minimizer $f_0 \in \mathcal{F}_{MC}$,*

$$\mathcal{H}(f) \geq \mathcal{H}(f_0) \quad \text{for all } f \in \mathcal{F}_{MC},$$

where the constraint set is defined as

$$\mathcal{F}_{MC} := \left\{ f \in L^1_+(\mathbb{R}^6) \mid \iint f \, dv \, dx + \mathcal{C}(f) = M, E_{\text{kin}}(f) < \infty \right\}.$$

In both cases the parameter $M > 0$ is a prescribed positive number. In order to obtain solutions to these problems we make the following assumptions on Φ :

ASSUMPTIONS ON Φ . Let $\Phi \in C^1([0, \infty[)$ with $\Phi(0) = 0 = \Phi'(0)$, and

($\Phi 1$) Φ is strictly convex,

($\Phi 2$) $\Phi(f) \geq C f^{1+1/k}$ for $f \geq 0$ large,

where $0 < k < 3/2$ for Version 1,

$0 < k < 7/2$ for Version 2.

($\Phi 3$) In addition for Version 1

$\Phi(f) \leq C f^{1+1/k'}$ for $f \geq 0$ small, where $0 < k' < 3/2$.

A typical function Φ is

$$\Phi(f) = \frac{k}{k+1} f^{1+1/k}, \quad f \geq 0. \quad (2.6)$$

The first version covers the parameter range $0 < k < 3/2$, the second one covers $0 < k < 7/2$, indeed, with some technical extra effort also the limiting case $k = 7/2$ can be covered, cf. [48]. The main advantage of the first approach, which covers a smaller range of the polytropic steady states, is that it can be attacked via a reduction procedure, and this reduction procedure brings out a relation between the stability problem for the Vlasov–Poisson system, i.e., for a self-gravitating collisionless gas, and the one for a self-gravitating perfect fluid as described by the Euler–Poisson system.

As will be seen below, the potential energy is finite for states in the constraint sets. The major step in the stability analysis is to prove the following theorem.

THEOREM 2.1. *Consider Version 1 of the variational problem under the above assumptions on Φ . Then the energy-Casimir functional $\mathcal{H}_C$ is bounded from below on $\mathcal{F}_M$ with $h_M := \inf_{\mathcal{F}_M} \mathcal{H}_C < 0$. Let $(f_j) \subset \mathcal{F}_M$ be a minimizing sequence of $\mathcal{H}_C$, i.e., $\mathcal{H}_C(f_j) \rightarrow h_M$. Then there exists a function $f_0 \in \mathcal{F}_M$, a subsequence, again denoted by (f_j) and a sequence $(a_j) \subset \mathbb{R}^3$ of shift vectors such that*

$$\begin{aligned} T^{a_j} f_j &:= f_j(\cdot + a_j, \cdot) \rightharpoonup f_0 && \text{weakly in } L^{1+1/k}(\mathbb{R}^6), j \rightarrow \infty, \\ T^{a_j} \nabla U_{f_j} &= \nabla U_{f_j}(\cdot + a_j) \rightarrow \nabla U_{f_0} && \text{strongly in } L^2(\mathbb{R}^3), j \rightarrow \infty. \end{aligned}$$

The state f_0 minimizes the energy-Casimir functional: $\mathcal{H}_C(f_0) = h_M$.

The analogous assertions hold for Version 2 of the variational problem, with $\mathcal{H}_C$ replaced by $\mathcal{H}$ and $\mathcal{F}_M$ by $\mathcal{F}_{MC}$.

Since the functionals under consideration are invariant under spatial translations a trivial minimizing sequence is obtained by shifting a given minimizer in space. If for example it is shifted off to infinity no subsequence can tend weakly to a minimizer, unless one moves with the sequence. Hence the spatial shifts in the theorem arise from the physical properties of the problem.

In Section 2.6 stability of the state f_0 will follow quite easily from the theorem. The point is that in the Taylor expansion (2.5) the negative definite part, i.e., the L^2 difference of the gravitational fields, converges to zero along minimizing sequences. Hence it is essential that the latter is part of Theorem 2.1 – the mere fact that f_0 be a minimizer is by itself not sufficient for stability.

The main difficulty of the proof of Theorem 2.1 is seen from the following sketch. To obtain a lower bound for the functional on the constraint set is easy, and by Assumption $(\Phi 2)$ minimizing sequences can be seen to be bounded in $L^{1+1/k}$. Hence such a sequence has a weakly convergent subsequence, cf. [74], Section 2.18. The weak limit f_0 is the candidate for the minimizer, and one has to pass the limit into the various functionals. This is easy for the kinetic energy, the latter being linear. The Casimir functional is convex due to Assumption $(\Phi 1)$, and so one can use Mazur's lemma for the same purpose, cf. [74],

Section 2.13. The difficult part is the potential energy, for which one has to prove that the induced gravitational fields converge strongly in L^2 . Since the latter do not depend directly on the phase space density f but only on the induced spatial density ρ_f the state space $\mathcal{F}_M$ seems inappropriate for the latter problem. This is the mathematical motivation for passing to a reduced functional which is defined on a suitable set of spatial densities. The reduction procedure is explained in the next section. Then we turn to the proof of Theorem 2.1 for the case of Version 1. For Version 2 reduction does not work and the necessary additional arguments are discussed in Section 2.4. The pay-off of reduction in terms of stability results for the Euler–Poisson system is discussed in Section 2.7.

2.2. Reduction

In this section we consider Version 1 of the variational problem with Φ satisfying the assumptions above; the results below are based on [97]. The aim is to factor out the v -dependence and obtain a reduced variational problem in terms of spatial densities. For $r \geq 0$ let

$$\mathcal{G}_r := \left\{ g \in L^1_+(\mathbb{R}^3) \mid \int \left(\frac{1}{2} |v|^2 g(v) + \Phi(g(v)) \right) dv < \infty, \int g(v) dv = r \right\} \quad (2.7)$$

and

$$\Psi(r) := \inf_{g \in \mathcal{G}_r} \int \left(\frac{1}{2} |v|^2 g(v) + \Phi(g(v)) \right) dv. \quad (2.8)$$

In addition to the variational problem of minimizing $\mathcal{H}_C$ over the set $\mathcal{F}_M$ we consider the problem of minimizing the functional

$$\mathcal{H}_r(\rho) := \int \Psi(\rho(x)) dx + E_{\text{pot}}(\rho) \quad (2.9)$$

over the set

$$\mathcal{R}_M := \left\{ \rho \in L^1_+(\mathbb{R}^3) \mid \int \Psi(\rho(x)) dx < \infty, \int \rho(x) dx = M \right\}; \quad (2.10)$$

it will be seen further that the potential energy $E_{\text{pot}}(\rho)$ which is defined in the obvious way is finite for states in this constraint set. The topic of the present section is the relation between the minimizers of $\mathcal{H}_C$ and $\mathcal{H}_r$. The following remark should convince the reader that the construction above is indeed a very natural one.

REMARK. Consider the intermediate functional

$$\mathcal{P}(\rho) := \inf_{f \in \mathcal{F}_\rho} \iint \left(\frac{1}{2} |v|^2 f(x, v) + \Phi(f(x, v)) \right) dv dx,$$

where for $\rho \in \mathcal{R}_M$,

$$\mathcal{F}_\rho := \{f \in \mathcal{F}_M \mid \rho_f = \rho\}.$$

Clearly, for $\rho = \rho_f$ with $f \in \mathcal{F}_M$,

$$\begin{aligned} \mathcal{C}(f) + E_{\text{kin}}(f) &\geq \inf_{\tilde{f} \in \mathcal{F}_\rho} (\mathcal{C}(\tilde{f}) + E_{\text{kin}}(\tilde{f})) \\ &\geq \inf_{\tilde{f} \in \mathcal{F}_\rho} \int \left[\inf_{g \in \mathcal{G}_{\rho(x)}} \int \left(\frac{1}{2} |v|^2 g(v) + \Phi(g(v)) \right) dv \right] dx \\ &= \int \left[\inf_{g \in \mathcal{G}_{\rho(x)}} \int \left(\frac{1}{2} |v|^2 g(v) + \Phi(g(v)) \right) dv \right] dx \\ &= \int \Psi(\rho(x)) dx. \end{aligned} \tag{2.11}$$

This shows that

$$\mathcal{H}_C(f) \geq \mathcal{P}(\rho_f) + E_{\text{pot}}(\rho_f) \geq \int \Psi(\rho_f(x)) dx + E_{\text{pot}}(\rho_f) = \mathcal{H}_r(\rho_f),$$

and it will be seen below that equality holds for minimizers. The functional $\mathcal{P}(\rho)$ is obtained by minimizing the positive contribution to $\mathcal{H}_C$, which also happens to be the part depending on phase space densities f directly, over all f 's which generate a given spatial density ρ . Then in a second step one minimizes for each point x over all functions $g = g(v)$ the integral of which has the value $\rho(x)$.

These constructions owe much to [112] where they appear for the special case $\Phi(f) = f^{1+1/k}$ in a spherically symmetric situation. The main result of the present section is the following theorem.

THEOREM 2.2. (a) For every function $f \in \mathcal{F}_M$,

$$\mathcal{H}_C(f) \geq \mathcal{H}_r(\rho_f),$$

and if $f = f_0$ is a minimizer of $\mathcal{H}_C$ over $\mathcal{F}_M$ then equality holds.

(b) Let $\rho_0 \in \mathcal{R}_M$ be a minimizer of $\mathcal{H}_r$ with induced potential U_0 . Then there exists a Lagrange multiplier $E_0 \in \mathbb{R}$ such that a.e.,

$$\rho_0 = \begin{cases} (\Psi')^{-1}(E_0 - U_0), & U_0 < E_0, \\ 0, & U_0 \geq E_0. \end{cases} \tag{2.12}$$

With the particle energy E defined as in (2.1) the function

$$f_0 := \begin{cases} (\Phi')^{-1}(E_0 - E), & E < E_0, \\ 0, & E \geq E_0, \end{cases}$$

is a minimizer of $\mathcal{H}_C$ in $\mathcal{F}_M$.

(c) Assume that $\mathcal{H}_r$ has a minimizer in $\mathcal{R}_M$. If $f_0 \in \mathcal{F}_M$ is a minimizer of $\mathcal{H}_C$ then $\rho_0 := \rho_{f_0} \in \mathcal{R}_M$ is a minimizer of $\mathcal{H}_r$, this map is one-to-one and onto between the sets of minimizers of $\mathcal{H}_C$ in $\mathcal{F}_M$ and of $\mathcal{H}_r$ in $\mathcal{R}_M$ respectively, and is the inverse of the map $\rho_0 \mapsto f_0$ described in (b).

In the next section we show that the reduced functional $\mathcal{H}_r$ does have a minimizer, and then the theorem guarantees that we recover all minimizers of $\mathcal{H}_C$ in $\mathcal{F}_M$ by “lifting” the ones of $\mathcal{H}_r$ as described in (b).

The above relation between Φ and Ψ arises in a natural way, but it can be made more explicit. Denote the Legendre transform of a function $h : \mathbb{R} \rightarrow]-\infty, \infty]$ by

$$\bar{h}(\lambda) := \sup_{r \in \mathbb{R}} (\lambda r - h(r)).$$

LEMMA 2.3. Let Ψ be defined by (2.7), (2.8), and extend both Φ and Ψ by $+\infty$ to the interval $]-\infty, 0[$.

(a) For $\lambda \in \mathbb{R}$,

$$\bar{\Psi}(\lambda) = \int \bar{\Phi}\left(\lambda - \frac{1}{2}|v|^2\right) dv,$$

and in particular, $\bar{\Phi}(\lambda) = 0 = \bar{\Psi}(\lambda)$ for $\lambda < 0$.

(b) $\Psi \in C^1([0, \infty[)$ is strictly convex, and $\Psi(0) = \Psi'(0) = 0$.

(c) With positive constants C which depend on Φ and M ,
 $\Psi(\rho) \geq C\rho^{1+1/n}$ for $\rho \geq 0$ large, where $n := k + 3/2$, and
 $\Psi(\rho) \leq C\rho^{1+1/n'}$ for $\rho \geq 0$ small, where $n' := k' + 3/2$.

PROOF. By definition,

$$\begin{aligned} \bar{\Psi}(\lambda) &= \sup_{r \geq 0} \left[\lambda r - \inf_{g \in \mathcal{G}_r} \int \left(\frac{1}{2}|v|^2 g(v) + \Phi(g(v)) \right) dv \right] \\ &= \sup_{r \geq 0} \sup_{g \in \mathcal{G}_r} \int \left[\left(\lambda - \frac{1}{2}|v|^2 \right) g(v) - \Phi(g(v)) \right] dv \\ &= \sup_{g \in L_+^1(\mathbb{R}^3)} \int \left[\left(\lambda - \frac{1}{2}|v|^2 \right) g(v) - \Phi(g(v)) \right] dv \\ &= \int \sup_{y \geq 0} \left[\left(\lambda - \frac{1}{2}|v|^2 \right) y - \Phi(y) \right] dv = \int \bar{\Phi}\left(\lambda - \frac{1}{2}|v|^2\right) dv. \end{aligned}$$

As to the last-but-one equality, observe that both sides are obviously zero for $\lambda \leq 0$. If $\lambda > 0$ then for any $g \in L_+^1(\mathbb{R}^3)$,

$$\int \left[\left(\lambda - \frac{1}{2}|v|^2 \right) g(v) - \Phi(g(v)) \right] dv \leq \int \sup_{y \geq 0} \left[\left(\lambda - \frac{1}{2}|v|^2 \right) y - \Phi(y) \right] dv.$$

If $|v| \geq \sqrt{2\lambda}$ then $\sup_{y \geq 0} [\cdots] = 0$, and for $|v| < \sqrt{2\lambda}$ the supremum of the term in brackets is attained at $y = y_v := (\Phi')^{-1}(\lambda - \frac{1}{2}|v|^2)$. Thus with

$$g_0(v) := \begin{cases} y_v, & |v| < \sqrt{2\lambda}, \\ 0, & |v| \geq \sqrt{2\lambda}, \end{cases}$$

we have

$$\begin{aligned} & \int \sup_{y \geq 0} \left[\left(\lambda - \frac{1}{2}|v|^2 \right) y - \Phi(y) \right] dv \\ &= \int \left[\left(\lambda - \frac{1}{2}|v|^2 \right) g_0(v) - \Phi(g_0(v)) \right] dv \\ &\leq \sup_{g \in L^1_+(\mathbb{R}^3)} \int \left[\left(\lambda - \frac{1}{2}|v|^2 \right) g(v) - \Phi(g(v)) \right] dv, \end{aligned}$$

and part (a) is established.

Since Φ is strictly convex and lower semicontinuous as a function on $\mathbb{R}$ with $\lim_{|f| \rightarrow \infty} \Phi(f)/|f| \rightarrow \infty$, $\bar{\Phi} \in C^1(\mathbb{R})$, cf. [84, Prop. 2.4]. Obviously, $\bar{\Phi}(\lambda) = 0$ for $\lambda \leq 0$, in particular, $(\bar{\Phi})'(0) = 0$. Also, $(\bar{\Phi})'$ is strictly increasing on $[0, \infty[$ since Φ' is strictly increasing on $[0, \infty[$ with range $[0, \infty[$. Since for $|\lambda| < \lambda_0$ with $\lambda_0 > 0$ fixed the integral in the formula for $\bar{\Psi}$ extends over a compact set we may differentiate under the integral sign to conclude that $\bar{\Psi} \in C^1(\mathbb{R})$ with derivative strictly increasing on $[0, \infty[$. This in turn implies the assertion of part (b).

Part (c) follows with (a) and the definition of the Legendre transform. $\square$

We now prove Theorem 2.1.

PROOF OF THEOREM 2.2. We start by proving

The Euler–Lagrange equation for the reduced problem. Let $\rho_0 \in \mathcal{R}_M$ be a minimizer with induced potential U_0 . For $\varepsilon > 0$ define

$$S_\varepsilon := \left\{ x \in \mathbb{R}^3 \mid \varepsilon \leq \rho_0(x) \leq \frac{1}{\varepsilon} \right\};$$

think of ρ_0 as a pointwise defined representative of the minimizer. For a test function $w \in L^\infty(\mathbb{R}^3)$ which has compact support and is nonnegative on $\mathbb{R}^3 \setminus S_\varepsilon$ define for $\tau \geq 0$ small,

$$\rho_\tau := \rho_0 + \tau w - \tau \frac{\int w \, dy}{\text{vol } S_\varepsilon} \mathbb{1}_{S_\varepsilon}.$$

Then $\rho_\tau \geq 0$ and $\int \rho_\tau = M$ so that $\rho_\tau \in \mathcal{R}_M$ for $\tau \geq 0$ small. Since ρ_0 is a minimizer of $\mathcal{H}_r$,

$$0 \leq \mathcal{H}_r(\rho_\tau) - \mathcal{H}_r(\rho_0) = \tau \int (\Psi'(\rho_0) + U_0) \left(w - \frac{\int w \, dy}{\text{vol } S_\varepsilon} \mathbb{1}_{S_\varepsilon} \right) dx + o(\tau).$$

Hence the coefficient of τ in this estimate must be nonnegative, which we can rewrite in the form

$$\int \left[\Psi'(\rho_0) + U_0 - \frac{1}{\text{vol } S_\varepsilon} \left(\int_{S_\varepsilon} (\Psi'(\rho_0) + U_0) \, dy \right) \right] w \, dx \geq 0.$$

This holds for all test functions w as specified above, and hence $\Psi'(\rho_0) + U_0 = E_\varepsilon$ on S_ε and $\Psi'(\rho_0) + U_0 \geq E_\varepsilon$ on $\mathbb{R}^3 \setminus S_\varepsilon$ for all $\varepsilon > 0$ small enough. Here E_ε is a constant which by the first relation must be independent of ε , and taking $\varepsilon \rightarrow 0$ proves the relation between ρ_0 and U_0 in part (b).

The inequality in part (a) was established as part of the remark before Theorem 2.2.

An intermediate assertion. We claim that if $f \in \mathcal{F}_M$ is such that up to sets of measure zero,

$$\begin{cases} \Phi'(f) = E_0 - E > 0, & \text{where } f > 0, \\ E_0 - E \leq 0, & \text{where } f = 0, \end{cases} \quad (2.13)$$

with $E := \frac{1}{2}|v|^2 + U_f(x)$ and E_0 a constant, then equality holds in (a). To prove this, observe that since Φ is convex, we have for a.e. $x \in \mathbb{R}^3$ and every $g \in \mathcal{G}_{\rho_f(x)}$,

$$\begin{aligned} \frac{1}{2}|v|^2 g(v) + \Phi(g(v)) &\geq \frac{1}{2}|v|^2 f(x, v) + \Phi(f(x, v)) \\ &\quad + \left(\frac{1}{2}|v|^2 + \Phi'(f(x, v)) \right) (g(v) - f(x, v)) \quad \text{a.e.} \end{aligned}$$

Now by (2.13),

$$\begin{aligned} &\int \left(\frac{1}{2}|v|^2 + \Phi'(f) \right) (g - f) \, dv \\ &= \int_{\{f>0\}} \cdots + \int_{\{f=0\}} \cdots \\ &= (E_0 - U_f(x)) \int_{\{f>0\}} (g - f) \, dv + \int_{\{f=0\}} \frac{1}{2}|v|^2 g \, dv \\ &= -(E_0 - U_f(x)) \int_{\{f=0\}} (g - f) \, dv + \int_{\{f=0\}} \frac{1}{2}|v|^2 g \, dv \\ &= \int_{\{f=0\}} (E - E_0) g \, dv \geq 0; \end{aligned}$$

observe that $g \geq 0$ and $\int (g - f) dv = 0$. Hence

$$\begin{aligned} \Psi(\rho_f(x)) &\geq \int \left(\frac{1}{2} |v|^2 f + \Phi(f) \right) dv \\ &\geq \inf_{g \in \mathcal{G}_{\rho_f(x)}} \int \left(\frac{1}{2} |v|^2 g + \Phi(g) \right) dv = \Psi(\rho_f(x)) \quad \text{a.e.}, \end{aligned}$$

and the proof of the intermediate assertion is complete.

Proof of the equality assertion in (a). If $f_0 \in \mathcal{F}_M$ is a minimizer of $\mathcal{H}_C$ then the Euler–Lagrange equation of the minimization problem implies that (2.13) holds for some Lagrange multiplier E_0 ; the proof is essentially the same as for the reduced problem above, cf. also Theorem 2.6. Thus equality holds in (a) by the intermediate assertion, and the proof of part (a) is complete.

Proof of the remaining part of (b). Let f_0 be defined as in (b). Then up to sets of measure zero,

$$\begin{aligned} \int f_0(x, v) dv &= \int_{|v| \leq \sqrt{2(E_0 - U_0(x))}} (\Phi')^{-1} \left(E_0 - U_0(x) - \frac{1}{2} |v|^2 \right) dv \\ &= (\bar{\Psi})'(E_0 - U_0(x)) = (\Psi')^{-1}(E_0 - U_0(x)) = \rho_0(x), \end{aligned}$$

where $U_0(x) < E_0$, and both sides are zero where $U_0(x) \geq E_0$. Thus $\rho_0 = \rho_{f_0}$, in particular, $f_0 \in \mathcal{F}_M$. By definition, f_0 satisfies the relation (2.13) and thus by our intermediate assertion $\mathcal{H}_C(f_0) = \mathcal{H}_r(\rho_0)$. Therefore again by part (a),

$$\mathcal{H}_C(f) \geq \mathcal{H}_r(\rho_f) \geq \mathcal{H}_r(\rho_0) = \mathcal{H}_C(f_0), \quad f \in \mathcal{F}_M,$$

so that f_0 is a minimizer of $\mathcal{H}_C$, and the proof of part (b) is complete.

Proof of part (c). Assume that $\mathcal{H}_r$ has a minimizer $\rho_0 \in \mathcal{R}_M$ and define f_0 as above. Then part (a), the fact that each $\rho \in \mathcal{R}_M$ can be written as $\rho = \rho_f$ for some $f \in \mathcal{F}_M$, and our intermediate assertion imply that

$$\begin{aligned} \inf_{f \in \mathcal{F}_M} \mathcal{H}_C(f) &\geq \inf_{f \in \mathcal{F}_M} \mathcal{H}_r(\rho_f) = \inf_{\rho \in \mathcal{R}_M} \mathcal{H}_r(\rho) \\ &= \mathcal{H}_r(\rho_0) = \mathcal{H}_C(f_0) \geq \inf_{f \in \mathcal{F}_M} \mathcal{H}_C(f). \end{aligned} \tag{2.14}$$

Now take any minimizer $g_0 \in \mathcal{F}_M$ of $\mathcal{H}_C$. Then by (2.14) and part (a),

$$\inf_{\rho \in \mathcal{R}_M} \mathcal{H}_r(\rho) = \inf_{f \in \mathcal{F}_M} \mathcal{H}_C(f) = \mathcal{H}_C(g_0) = \mathcal{H}_r(\rho_{g_0}),$$

that is, $\rho_{g_0} \in \mathcal{R}_M$ minimizes $\mathcal{H}_r$, and the proof of part (c) is complete. $\square$

2.3. Existence of minimizers via the reduced problem

First the reduced variational problem is studied in its own right under the following assumptions on the function Ψ .

ASSUMPTIONS ON Ψ . Let $\Psi \in C^1([0, \infty[)$ with $\Psi(0) = 0 = \Psi'(0)$, and

($\Psi 1$) Ψ is strictly convex,

($\Psi 2$) $\Psi(\rho) \geq C\rho^{1+1/n}$ for $\rho \geq 0$ large, with $0 < n < 3$,

($\Psi 3$) $\Psi(\rho) \leq C\rho^{1+1/n'}$ for $\rho \geq 0$ small, with $0 < n' < 3$.

We shall prove the following central result:

THEOREM 2.4. *The functional $\mathcal{H}_r$ is bounded from below on $\mathcal{R}_M$. Let $(\rho_j) \subset \mathcal{R}_M$ be a minimizing sequence of $\mathcal{H}_r$. Then there exists a sequence of shift vectors $(a_j) \subset \mathbb{R}^3$ and a subsequence, again denoted by (ρ_j) , such that*

$$T^{a_j} \rho_j := \rho_j(\cdot + a_j) \rightharpoonup \rho_0 \quad \text{weakly in } L^{1+1/n}(\mathbb{R}^3), j \rightarrow \infty,$$

$$T^{a_j} \nabla U_{\rho_j} \rightarrow \nabla U_{\rho_0} \quad \text{strongly in } L^2(\mathbb{R}^3), j \rightarrow \infty,$$

and $\rho_0 \in \mathcal{R}_M$ is a minimizer of $\mathcal{H}_r$.

The main difficulty is to prove that the fields induced by a minimizing sequence converge strongly in L^2 . Such a compactness property holds if the sequence (ρ_j) remains concentrated. In view of the next section the corresponding result stated below is slightly more general than what is needed in the present section.

LEMMA 2.5. *Let $0 < n < 5$. Let $(\rho_j) \subset L^{1+1/n}_+(\mathbb{R}^3)$ be such that*

$$\begin{aligned} \rho_j &\rightharpoonup \rho_0 \quad \text{weakly in } L^{1+1/n}(\mathbb{R}^3), \\ \forall \varepsilon > 0 \exists R > 0: \quad \limsup_{j \rightarrow \infty} \int_{|x| \geq R} \rho_j(x) dx &< \varepsilon. \end{aligned} \tag{2.15}$$

Then $\nabla U_{\rho_j} \rightarrow \nabla U_{\rho_0}$ strongly in L^2 .

PROOF. Clearly, there exists a constant $m > 0$ such that for all sufficiently large $j \in \mathbb{N}$, $\int \rho_j \leq m$, and by weak convergence the limit ρ_0 is integrable as well. The sequence $\sigma_j := \rho_j - \rho_0$ converges weakly to 0 in $L^{1+1/n}$, $\int |\sigma_j| \leq 2m$, and (2.15) holds for $|\sigma_j|$ as well. We need to show that $\nabla U_{\sigma_j} \rightarrow 0$ strongly in L^2 which is equivalent to

$$I_j := \iint \frac{\sigma_j(x)\sigma_j(y)}{|x-y|} dy dx \rightarrow 0.$$

For $\delta > 0$ and $R > 0$ we split the integral above as follows:

$$I_j = I_{j,1} + I_{j,2} + I_{j,3},$$

where

$$\begin{aligned} |x - y| < \delta & \text{ for } I_{j,1}, & |x - y| \geq \delta \wedge (|x| \geq R \vee |y| \geq R) & \text{ for } I_{j,2}, \\ |x - y| \geq \delta \wedge |x| < R \wedge |y| < R & \text{ for } I_{j,3}. \end{aligned}$$

Since $2n/(n+1) + 2/(n+1) = 2$, Young's inequality [74], Section 4.2, implies that

$$|I_{j,1}| \leq C \|\sigma_j\|_{1+1/n}^2 \|\mathbb{1}_{B_\delta} \cdot |\cdot|^{-1}\|_{(n+1)/2} \leq C \delta^{(5-n)/(n+1)}.$$

Hence we can make $I_{j,1}$ as small as we wish, uniformly in j and independently of R , by choosing δ small. For $\delta > 0$ now fixed,

$$|I_{j,2}| \leq \frac{4m}{\delta} \int_{|x| > R} |\sigma_j(x)| dx,$$

which becomes small by (2.15), if we choose $R > 0$ accordingly. Finally by Hölder's inequality,

$$|I_{j,3}| = \left| \int \sigma_j(x) h_j(x) dx \right| \leq \|\sigma_j\|_{1+1/n} \|h_j\|_{1+n} \leq C \|h_j\|_{1+n},$$

where in a pointwise sense,

$$h_j(x) := \mathbb{1}_{B_R}(x) \int_{|x-y| \geq \delta} \mathbb{1}_{B_R}(y) \frac{1}{|x-y|} \sigma_j(y) dy \rightarrow 0$$

due to the weak convergence of σ_j and the fact that the test function against which σ_j is integrated here is in L^{1+n} . Since $|h_j| \leq \frac{2m}{\delta} \mathbb{1}_{B_R}$ uniformly in j Lebesgue's dominated convergence theorem implies that $h_j \rightarrow 0$ in L^{1+n} , and the proof is complete. $\square$

PROOF OF THEOREM 2.4. Constants denoted by C may only depend on M and Ψ and may change their value from line to line. The proof is split into a number of steps.

Step 1: Lower bound for $\mathcal{H}_r$ and weak convergence of minimizing sequences. By Lemma P2(b), interpolation, and $(\Psi 2)$,

$$\begin{aligned} -E_{\text{pot}}(\rho) & \leq C \|\rho\|_{6/5}^2 \leq C \|\rho\|_1^{(5-n)/3} \|\rho\|_{1+1/n}^{(n+1)/3} \\ & \leq C + C \left(\int \Psi(\rho) dx \right)^{n/3}, \quad \rho \in \mathcal{R}_M; \end{aligned}$$

note that $1 < 6/5 < 1 + 1/n$. Hence on $\mathcal{R}_M$

$$\mathcal{H}_r(\rho) \geq \int \Psi(\rho) dx - C - C \left(\int \Psi(\rho) dx \right)^{n/3}. \quad (2.16)$$

Since $n < 3$ this implies that $\mathcal{H}_r$ is bounded from below on $\mathcal{R}_M$,

$$h_M := \inf_{\mathcal{R}_M} \mathcal{H}_r > -\infty.$$

Let $(\rho_j) \subset \mathcal{R}_M$ be a minimizing sequence. By (2.16), $\int \Psi(\rho_j)$ is bounded, and by $(\Psi 2)$ and the fact that $\int \rho_j = M$, the minimizing sequence is bounded in $L^{1+1/n}(\mathbb{R}^3)$. Hence we can – after extracting a subsequence – assume that it converges weakly to some function $\rho_0 \in L^{1+1/n}(\mathbb{R}^3)$. By weak convergence, $\rho_0 \geq 0$ almost everywhere – if ρ_0 were strictly negative on some set S of positive, finite measure the test function $\sigma = \mathbb{1}_S$ would yield a contradiction.

The next two steps show that minimizing sequences remain concentrated and do not split into far apart pieces or spread out uniformly in space.

Step 2: Behavior under rescaling. For $\rho \in \mathcal{R}_M$ and $a, b > 0$ we define $\bar{\rho}(x) := a\rho(bx)$. Then

$$\begin{aligned} \int \bar{\rho} \, dx &= ab^{-3} \int \rho \, dx, & E_{\text{pot}}(\bar{\rho}) &= a^2 b^{-5} E_{\text{pot}}(\rho), \\ \int \Psi(\bar{\rho}) &= b^{-3} \int \Psi(a\rho) \, dx. \end{aligned}$$

First we fix a bounded and compactly supported function $\rho \in \mathcal{R}_M$ and choose $a = b^3$ so that $\bar{\rho} \in \mathcal{R}_M$ as well. By $(\Psi 3)$ and since $3/n' > 1$,

$$\mathcal{H}_r(\bar{\rho}) = b^{-3} \int \Psi(b^3 \rho) \, dx + b E_{\text{pot}}(\rho) \leq C b^{3/n'} + b E_{\text{pot}}(\rho) < 0$$

for b sufficiently small, and hence for $M > 0$,

$$h_M < 0. \tag{2.17}$$

Next we fix two masses $0 < \bar{M} \leq M$. If we take $a = 1$ and $b = (M/\bar{M})^{1/3} \geq 1$ then for $\rho \in \mathcal{R}_M$ and $\bar{\rho} \in \mathcal{R}_{\bar{M}}$ rescaled with these parameters,

$$\begin{aligned} \mathcal{H}_r(\bar{\rho}) &= b^{-3} \int \Psi(\rho) \, dx + b^{-5} E_{\text{pot}}(\rho) \\ &\geq b^{-5} \left(\int \Psi(\rho) \, dx + E_{\text{pot}}(\rho) \right) = \left(\frac{\bar{M}}{M} \right)^{5/3} \mathcal{H}_r(\rho). \end{aligned}$$

Since for the present choice of a and b the map $\rho \mapsto \bar{\rho}$ is one-to-one and onto between $\mathcal{R}_M$ and $\mathcal{R}_{\bar{M}}$ this estimate gives the following relation between the infima of our functional for different mass constraints:

$$h_{\bar{M}} \geq \left(\frac{\bar{M}}{M} \right)^{5/3} h_M, \quad 0 < \bar{M} \leq M. \tag{2.18}$$

Step 3: Spherically symmetric minimizing sequences remain concentrated. In this step we prove the concentration property needed to apply Lemma 2.5, but to make things easier we consider for a moment spherically symmetric functions $\rho \in \mathcal{R}_M$, i.e., $\rho(x) = \rho(|x|)$. For any radius $R > 0$ we split ρ into the piece supported in the ball B_R and the rest, i.e.,

$$\rho = \rho_1 + \rho_2, \quad \rho_1(x) = 0 \text{ for } |x| > R, \quad \rho_2(x) = 0 \text{ for } |x| \leq R.$$

Clearly,

$$\mathcal{H}_r(\rho) = \mathcal{H}_r(\rho_1) + \mathcal{H}_r(\rho_2) - \int \frac{\rho_1(x)\rho_2(y)}{|x-y|} dx dy.$$

Due to spherical symmetry the potential energy of the interaction between the two pieces can be estimated as

$$\int \frac{\rho_1(x)\rho_2(y)}{|x-y|} dx dy = - \int U_{\rho_1} \rho_2 dx \leq \frac{(M-m)m}{R},$$

where $m = \int \rho_2$ is the mass outside the radius R which we want to make small along the minimizing sequence. We define

$$R_0 := -\frac{3}{5} \frac{M^2}{h_M} > 0$$

and use the scaling estimate (2.18) together with (2.17) and the fact that $\xi^{5/3} + (1-\xi)^{5/3} \leq 1 - \frac{5}{3}\xi(1-\xi)$ for $0 \leq \xi \leq 1$ to conclude that

$$\begin{aligned} \mathcal{H}_r(\rho) &\geq h_{M-m} + h_m - \frac{(M-m)m}{R} \\ &\geq \left[\left(1 - \frac{m}{M}\right)^{5/3} + \left(\frac{m}{M}\right)^{5/3} \right] h_M - \frac{(M-m)m}{R} \\ &\geq h_M + \left[\frac{1}{R_0} - \frac{1}{R} \right] (M-m)m. \end{aligned} \tag{2.19}$$

We claim that, if $R > R_0$, then for any spherically symmetric minimizing sequence $(\rho_j) \subset \mathcal{R}_M$ of $\mathcal{H}_r$, the following assertion holds, which is even a bit stronger than what is needed to apply lemma 2.5,

$$\lim_{j \rightarrow \infty} \int_{|x| \geq R} \rho_j(x) dx = 0. \tag{2.20}$$

Assume this assertion were false so that up to a subsequence,

$$\lim_{j \rightarrow \infty} \int_{|x| \geq R} \rho_j = m > 0.$$

Choose $R_j > R$ such that

$$m_j := \int_{|x| \geq R_j} \rho_j = \frac{1}{2} \int_{|x| \geq R} \rho_j.$$

By (2.19),

$$\mathcal{H}_r(\rho_j) \geq h_M + \left[\frac{1}{R_0} - \frac{1}{R_j} \right] (M - m_j) m_j \geq h_M + \left[\frac{1}{R_0} - \frac{1}{R} \right] (M - m_j) m_j,$$

and letting $j \rightarrow \infty$ leads to a contradiction, and equation (2.20) is proven.

For the weak limit ρ_0 of the minimizing sequence clearly

$$\text{supp } \rho_0 \subset B_{R_0}, \quad \int \rho_0 = M.$$

Step 4: Proof of Theorem 2.4 under the assumption of spherical symmetry. Given a minimizing sequence (ρ_j) we already know that up to a subsequence it converges weakly in $L^{1+1/n}$ to a nonnegative limit ρ_0 of mass M . The functional $\rho \mapsto \int \Psi(\rho) dx$ is convex by Assumption $(\Psi 1)$, so by Mazur's lemma [74], Section 2.13 and Fatou's lemma [74], Section 1.7

$$\int \Psi(\rho_0) dx \leq \limsup_{j \rightarrow \infty} \int \Psi(\rho_j) dx,$$

in particular, $\rho_0 \in \mathcal{R}_M$. If we assume in addition that the minimizing sequence is spherically symmetric then by Step 3 and Lemma 2.5, $E_{\text{pot}}(\rho_j) \rightarrow E_{\text{pot}}(\rho_0)$, and hence

$$\mathcal{H}_r(\rho_0) \leq \limsup_{j \rightarrow \infty} \mathcal{H}_r(\rho_j) = h_M$$

so that ρ_0 is a minimizer of $\mathcal{H}_r$ over the subset of spherically symmetric functions in $\mathcal{R}_M$.

The restriction to spherical symmetry would mean that stability would only hold against spherically symmetric perturbations. Fortunately, this restriction can be removed using a general result due to Burchard and Guo.

Step 5: Removing the symmetry assumption. To explain the result by Burchard and Guo we define for a given function $\rho \in L^1_+(\mathbb{R}^3)$ its spherically symmetric decreasing rearrangement ρ^* as the unique spherically symmetric, radially decreasing function with the property that for every $\tau \geq 0$ the sup-level-sets $\{x \in \mathbb{R}^3 \mid \rho(x) > \tau\}$ and $\{x \in \mathbb{R}^3 \mid \rho^*(x) > \tau\}$ have the same volume; the latter set is of course a ball about the origin whose radius is determined by the volume of the former. The integral $\int \Psi(\rho) dx$ does not change under such a rearrangement, while the potential energy can only decrease, and it does not decrease iff ρ is already spherically symmetric (with respect to some center of symmetry) and decreasing. These facts can be found in [74], Chapter 3. In particular, a minimizer must a posteriori be spherically symmetric.

Now let $(\rho_j) \subset \mathcal{R}_M$ be a not necessarily spherically symmetric minimizing sequence. Obviously, the sequence of spherically symmetric decreasing rearrangements (ρ_j^*) is again minimizing. Hence by the previous steps, up to a subsequence (ρ_j^*) converges weakly to a minimizer $\rho_0 = \rho_0^*$ and

$$\nabla U_{\rho_j^*} \rightarrow \nabla U_0 \quad \text{in } L^2, \quad \text{hence } \int \Psi(\rho_j^*) \rightarrow \int \Psi(\rho_0).$$

Moreover,

$$\begin{aligned} E_{\text{pot}}(\rho_j) &= \mathcal{H}_r(\rho_j) - \int \Psi(\rho_j) = \mathcal{H}_r(\rho_j) - \int \Psi(\rho_j^*) \\ &\rightarrow \mathcal{H}_r(\rho_0) - \int \Psi(\rho_0) = E_{\text{pot}}(\rho_0). \end{aligned}$$

In this situation the result of Burchard and Guo [16], Theorem 1, says that there exists a sequence $(a_j) \subset \mathbb{R}^3$ of shift vectors such that

$$T^{a_j} \nabla U_{\rho_j} = \nabla U_{\rho_j}(\cdot + a_j) \rightarrow \nabla U_0 \quad \text{in } L^2.$$

Hence we can repeat the arguments of Step 4 for the sequence $(T^{a_j} \rho_j)$, which is again minimizing, and the proof of Theorem 2.4 is complete. $\square$

The proof of the result by Burchard and Guo is by no means easy, and it is possible to obtain stability against general perturbations without resorting to it, cf. [48,96,97]. Since this general result may be useful for other problems of this nature we wanted to mention and exploit it here. On the other hand, Version 2 of the variational problem does not lend itself to a reduction mechanism like Version 1. Hence the result by Burchard and Guo does not apply, and we will show in the next section how to handle the concentration problem directly in the nonsymmetric situation. We also refer to [53] for an account of the result by Burchard and Guo and its relation to stability problems.

Theorem 2.4 implies the result that we were originally interested in.

PROOF OF THEOREM 2.1 FOR VERSION 1. By Lemma 2.3 we see that if Φ satisfies the assumptions $(\Phi 1)$, $(\Phi 2)$, $(\Phi 3)$ then the function Ψ defined by (2.8) satisfies the assumptions $(\Psi 1)$, $(\Psi 2)$, $(\Psi 3)$, where the parameters k and n are related by $n = k + 3/2$, with the same relation holding for the primed parameters. Theorem 2.2 connects the original and the reduced variational problem in the appropriate way to derive Theorem 2.1 from Theorem 2.4: Firstly, $\mathcal{H}_C$ is bounded from below on $\mathcal{F}_M$ since this is true for $\mathcal{H}_r$ on $\mathcal{R}_M$. Let $(f_j) \subset \mathcal{F}_M$ be a minimizing sequence for $\mathcal{H}_C$. By Theorem 2.2, $(\rho_{f_j}) \subset \mathcal{R}_M$ is a minimizing sequence for $\mathcal{H}_r$. Again by Theorem 2.2 we can lift the minimizer ρ_0 of $\mathcal{H}_r$ obtained in Theorem 2.4 to a minimizer f_0 of $\mathcal{H}_C$. The properly shifted fields converge strongly in L^2 to ∇U_{f_0} . Hence after extracting a subsequence the Casimir functional as well as the kinetic energy converge along $(T^{a_j} f_j)$, and this sequence converges weakly in $L^{1+1/k}$ to the minimizer f_0 . $\square$

Notice that the weak convergence of a subsequence of $(T^{a_j} f_j)$ to the minimizer f_0 , which was derived after the existence of the minimizer was established, will play a role in the stability analysis in Section 2.6.

2.4. Existence of minimizers – the direct approach

In this section we prove Theorem 2.1 for Version 2 of our variational problem, i.e., we minimize the energy functional $\mathcal{H}$ under the mass-Casimir constraint implemented in the constraint set $\mathcal{F}_{MC}$. In the reduction procedure employed above the kinetic energy and the Casimir functional were reduced into a new functional acting on spatial densities ρ . But in Version 2 of the variational problem the former two functionals appear in different places, namely as part of the functional to be minimized and in the constraint respectively. Hence reduction in the above sense does not apply, and a direct argument is given. This necessarily also shows how the use of the nontrivial result by Burchard and Guo for removing the symmetry assumption can be avoided.

PROOF OF THEOREM 2.1 FOR VERSION 2. Constants denoted by C may only depend on M and Φ and may change their value from line to line. The growth parameter k in the assumptions on Φ satisfies

$$0 < k < \frac{7}{2}, \quad \text{hence } \frac{3}{2} < n := k + \frac{3}{2} < 5 \quad \text{and} \quad 1 + \frac{1}{n} > \frac{6}{5}.$$

The proof is again split into a number of steps, similar to Version 1.

Step 1: Lower bound for $\mathcal{H}$ and bounds on minimizing sequences. By the assumptions on Φ , Lemma P2, Lemma 1.8 of Chapter 1, and interpolation the following estimates hold for any $f \in \mathcal{F}_{MC}$:

$$\begin{aligned} \|f\|_1 + \|f\|_{1+1/k} &\leq C, \\ \|\rho_f\|_{1+1/n} &\leq C \|f\|_{1+1/k}^{(k+1)/(n+1)} E_{\text{kin}}(f)^{3/(2k+5)} \leq C E_{\text{kin}}(f)^{3/(2(n+1))}, \\ -E_{\text{pot}}(f) &\leq C \|\rho_f\|_{6/5}^2 \leq C \|\rho_f\|_1^{(5-n)/3} \|\rho_f\|_{1+1/n}^{(n+1)/3} \leq C E_{\text{kin}}(f)^{1/2}. \end{aligned}$$

Hence the total energy $\mathcal{H}$ is bounded from below on $\mathcal{F}_{MC}$,

$$\mathcal{H}(f) \geq E_{\text{kin}}(f) - C E_{\text{kin}}(f)^{1/2} \quad \text{for } f \in \mathcal{F}_{MC}, \quad h_M := \inf_{\mathcal{F}_{MC}} \mathcal{H} > -\infty,$$

and E_{kin} together with the quantities estimated above are bounded along minimizing sequences of $\mathcal{H}$ in $\mathcal{F}_{MC}$.

The observation that concentration implies compactness made in Lemma 2.5 is going to be used again in the present situation, and we turn to the investigation of the concentration properties of the energy functional under the mass-Casimir constraint.

Step 2: Behavior under rescaling. Given any function f , we define a rescaled function $\bar{f}(x, v) = f(ax, bv)$, where $a, b > 0$; as opposed to Version 1 we do not scale the dependent variable but only its arguments. Then

$$\iint (\bar{f} + \Phi(\bar{f})) dv dx = (ab)^{-3} \iint (f + \Phi(f)) dv dx \quad (2.21)$$

i.e., $f \in \mathcal{F}_{MC}$ iff $\bar{f} \in \mathcal{F}_{\bar{M}C}$ where $\bar{M} := (ab)^{-3}M$. The kinetic and potential energy scale as follows:

$$E_{\text{kin}}(\bar{f}) = a^{-3}b^{-5}E_{\text{kin}}(f), \quad E_{\text{pot}}(\bar{f}) = a^{-5}b^{-6}E_{\text{pot}}(f).$$

If $f \in \mathcal{F}_{MC}$ and $b = a^{-1}$ then $\bar{f} \in \mathcal{F}_{MC}$ and

$$\mathcal{H}(\bar{f}) = a^2 E_{\text{kin}}(f) + a E_{\text{pot}}(f) < 0$$

for $a > 0$ sufficiently small, since $E_{\text{pot}}(f) < 0$. Hence for all $M > 0$,

$$h_M < 0. \quad (2.22)$$

Next we choose a and b such that $a^{-3}b^{-5} = a^{-5}b^{-6}$, i.e., $b = a^{-2}$. Then

$$\mathcal{H}(\bar{f}) = a^7 \mathcal{H}(f), \quad (2.23)$$

and since $a = (\bar{M}/M)^{1/3}$ and the mapping $\mathcal{F}_{MC} \rightarrow \mathcal{F}_{\bar{M}C}$, $f \mapsto \bar{f}$ is one-to-one and onto this shows that for all M , $\bar{M} > 0$,

$$h_{\bar{M}} = \left(\frac{\bar{M}}{M}\right)^{7/3} h_M. \quad (2.24)$$

Step 3: Minimizing sequences do not vanish. In the nonsymmetric case we cannot establish a result like equation (2.15) as easily as in the spherically symmetric situation. As a first step we show that along any minimizing sequence some minimal mass must remain in a sufficiently large ball. This is precisely the point where we have to allow spatial shifts: We cannot expect this nonvanishing property to hold unless we move with the sequence. Our assertion is that for any minimizing sequence $(f_j) \subset \mathcal{F}_{MC}$ of $\mathcal{H}$ there exist a sequence $(a_j) \subset \mathbb{R}^3$ and $m_0 > 0$, $R_0 > 0$ such that

$$\int_{a_j + B_{R_0}} \rho_j dx \geq m_0 \quad (2.25)$$

for all sufficiently large $j \in \mathbb{N}$, where $\rho_j := \rho_{f_j}$. To see this we split for $R > 1$,

$$-E_{\text{pot}}(f_j) = \frac{1}{2} \iint \frac{\rho_j(x)\rho_j(y)}{|x-y|} dy dx = I_1 + I_2 + I_3,$$

where

$$|x - y| < \frac{1}{R} \quad \text{for } I_1, \quad \frac{1}{R} \leq |x - y| \leq R \quad \text{for } I_2, \quad |x - y| > R \quad \text{for } I_3.$$

Since (ρ_j) is bounded in $L^1(\mathbb{R}^3)$ and in $L^{1+1/n}(\mathbb{R}^3)$ by Step 1,

$$\begin{aligned} I_1 &\leq \|\rho_j\|_{1+1/n}^2 \|\mathbb{1}_{B_{1/R}} 1/|\cdot|\|_{(n+1)/2} \leq CR^{-(5-n)/(n+1)}, \\ I_2 &\leq R \iint_{|x-y|<R} \rho_j(x) \rho_j(y) \, dx \, dy \leq RC \sup_{y \in \mathbb{R}^3} \int_{y+B_R} \rho_j(x) \, dx, \\ I_3 &\leq \frac{1}{R} \iint \rho_j(x) \rho_j(y) \, dx \, dy \leq CR^{-1}; \end{aligned}$$

for the first estimate we used Young's inequality [74], Section 4.2. Since (f_j) is minimizing and $h_M < 0$ we have, for any $R > 1$,

$$\frac{h_M}{2} > \mathcal{H}(f_j) \geq -I_1 - I_2 - I_3,$$

provided j is sufficiently large. Therefore,

$$\sup_{y \in \mathbb{R}^3} \int_{y+B_R} \rho_j \, dx \geq R^{-1} \left[-\frac{h_M}{2C} - R^{-1} - R^{-(5-n)/(n+1)} \right].$$

Since $h_M < 0$ the right-hand side of this estimate is positive for R sufficiently large, and the proof of equation (2.25) is complete.

Step 4: Nonvanishing, weakly convergent minimizing sequences remain concentrated. In this step we show that a minimizing sequence $(f_j) \subset \mathcal{F}_{MC}$ for $\mathcal{H}$ remains concentrated in the sense that equation (2.15) holds, provided that

$$\int_{B_{R_0}} \rho_j \, dx \geq m_0 \quad \text{and} \quad \rho_j \rightharpoonup \rho_0 \quad \text{weakly in } L^{1+1/n}(\mathbb{R}^3)$$

for some $R_0 > 0$ and $m_0 > 0$, where $\rho_j := \rho_{f_j}$. Notice that a minimizing sequence, if properly shifted in space, does not vanish by Step 3. Since the shifted minimizing sequence is again minimizing, the induced spatial densities do by Step 1 converge weakly as required after extracting a subsequence.

For $R > R_0$ we split f_j as follows:

$$f_j = f_j^1 + f_j^2 + f_j^3,$$

where

$$\begin{aligned} f_j^1(x, v) &= 0 \quad \text{for } |x| \geq R_0, \\ f_j^2(x, v) &= 0 \quad \text{for } |x| < R_0 \vee |x| > R, \\ f_j^3(x, v) &= 0 \quad \text{for } |x| \leq R. \end{aligned}$$

Then

$$\begin{aligned} \mathcal{H}(f_j) &= \mathcal{H}(f_j^1) + \mathcal{H}(f_j^2) + \mathcal{H}(f_j^3) \\ &\quad - \iint \frac{\rho_j^2(x)(\rho_j^1 + \rho_j^3)(y)}{|x - y|} dx dy - \iint \frac{\rho_j^1(x)\rho_j^3(y)}{|x - y|} dx dy \\ &=: \mathcal{H}(f_j^1) + \mathcal{H}(f_j^2) + \mathcal{H}(f_j^3) - I_1 - I_2, \end{aligned} \quad (2.26)$$

with obvious definitions for $\rho_j^1, \rho_j^2, \rho_j^3$. Since $\|\nabla U_{\rho_j^1 + \rho_j^3}\|_2$ is bounded by Step 1,

$$I_1 \leq C \|\nabla U_{\rho_j^2}\|_2 \leq C (\|\nabla U_{\rho_0^2}\|_2 + \|\nabla U_{\rho_j^2} - \nabla U_{\rho_0^2}\|_2).$$

For $R > 2R_0$ and $|x| \leq R_0, |y| > R$ we have $|x - y| \geq R/2$, and hence

$$I_2 \leq 2M^2 R^{-1}.$$

It is easy to show that $\xi^{7/3} + (1 - \xi)^{7/3} \leq 1 - \frac{7}{3}\xi(1 - \xi)$ for $\xi \in [0, 1]$. With equation (2.24) and obvious definitions of M_j^1, M_j^2, M_j^3 this implies that

$$\begin{aligned} \mathcal{H}(f_j^1) + \mathcal{H}(f_j^2) + \mathcal{H}(f_j^3) &\geq h_{M_j^1} + h_{M_j^2} + h_{M_j^3} \\ &= \left[\left(\frac{M_j^1}{M} \right)^{7/3} + \left(\frac{M_j^2}{M} \right)^{7/3} + \left(\frac{M_j^3}{M} \right)^{7/3} \right] h_M \\ &\geq \left[\left(\frac{M_j^1 + M_j^2}{M} \right)^{7/3} + \left(\frac{M_j^3}{M} \right)^{7/3} \right] h_M \\ &\geq \left[1 - \frac{7}{3} \frac{M_j^1 + M_j^2}{M} \frac{M_j^3}{M} \right] h_M \\ &\geq \left[1 - \frac{7}{3} \frac{m_0}{M^2} M_j^3 \right] h_M; \end{aligned}$$

in the last estimate we used the nonvanishing property. With (2.26) and the estimates for I_1 and I_2 this implies that

$$C_1 m_0 M_j^3 \leq \mathcal{H}(f_j) - h_M + C_2 [\|\nabla U_{\rho_0^2}\|_2 + \|\nabla U_{\rho_j^2} - \nabla U_{\rho_0^2}\|_2 + R^{-1}].$$

Here $R > 2R_0$ is so far arbitrary, and the constants C_1, C_2 are independent of R and R_0 . The first difference on the right-hand side converges to zero since the sequence (f_j) is minimizing. The first term in the bracket can be made as small as we wish by increasing R_0 ; notice that this does not affect the nonvanishing property. Choosing $R > 2R_0$ large makes the third term in the bracket small. For fixed $R_0 < R$ the middle term converges to zero by Lemma 2.5, since $\rho_j^2 \rightharpoonup \rho_0^2$ weakly in $L^{1+1/n}$ and these functions are supported in B_R . This shows that the sequence (f_j) satisfies the concentration property (2.15) as claimed.

Step 5: Proof of Theorem 2.1 for Version 2. Let (f_j) be a minimizing sequence and choose $(a_j) \subset \mathbb{R}^3$ according to (2.25). Since $\mathcal{H}$ is translation invariant $(T^{a_j} f_j)$ is again a minimizing sequence which by abuse of notation we denote by (f_j) . By Step 1, (f_j) is bounded in $L^{1+1/k}(\mathbb{R}^6)$. Thus there exists a weakly convergent subsequence, again denoted by (f_j) : $f_j \rightharpoonup f_0$. Clearly, $f_0 \geq 0$ a.e. Again by Step 1, $(E_{\text{kin}}(f_j))$ is bounded, and by weak convergence

$$E_{\text{kin}}(f_0) \leq \limsup_{j \rightarrow \infty} E_{\text{kin}}(f_j) < \infty.$$

By Step 1, $(\rho_j) = (\rho_{f_j})$ is bounded in $L^{1+1/n}(\mathbb{R}^3)$. After extracting a further subsequence

$$\rho_j \rightharpoonup \rho_0 = \rho_{f_0} \quad \text{weakly in } L^{1+1/n}(\mathbb{R}^3);$$

it is easy to see that the weak limit of the spatial densities induced by (f_j) is indeed the spatial density induced by the weak limit of (f_j) . By Step 4 and Lemma 2.5,

$$\nabla U_{\rho_j} \rightarrow \nabla U_0 \quad \text{strongly in } L^2(\mathbb{R}^3).$$

Hence $\mathcal{H}(f_0) \leq \lim_{j \rightarrow \infty} \mathcal{H}(f_j)$, and it remains to show that $\int f_0 + \mathcal{C}(f_0) = M$. By $(\Phi 2)$, Mazur's lemma, and Fatou's lemma

$$M_0 := \iint (f_0 + \Phi(f_0)) \, dv \, dx \leq \limsup_{j \rightarrow \infty} \iint (f_j + \Phi(f_j)) \, dv \, dx = M,$$

and $M_0 > 0$ since otherwise $f_0 = 0$ in contradiction to $\mathcal{H}(f_0) < 0$. Let

$$b := \left(\frac{M_0}{M} \right)^{2/3}, \quad a := b^{-1/2},$$

so that by (2.21), $\tilde{f}_0 \in \mathcal{F}_{MC}$. Then by (2.23),

$$h_M \leq \mathcal{H}(\tilde{f}_0) = a^7 \mathcal{H}(f_0) = \left(\frac{M}{M_0} \right)^{7/3} h_M,$$

which implies that $M_0 \geq M$. □

REMARK. Instead of the explicit arguments above one can also employ the concentration–compactness principle due to Lions [79], cf. [96].

2.5. Minimizers are steady states

Via the corresponding Euler–Lagrange identity the minimizers obtained by Theorem 2.1 are shown to be steady states of the Vlasov–Poisson system. Once the minimizers are identified as steady states some further properties are investigated. The minimizers obtained for the reduced variational problem in Theorem 2.4 turn out to be steady states of the Euler–Poisson system. This fact and the relation between steady states of the Vlasov–Poisson and of the Euler–Poisson system are postponed to Section 2.7.

THEOREM 2.6. *Let $f_0 \in \mathcal{F}_M$ be a minimizer of $\mathcal{H}_C$ with potential U_0 , and define the particle energy as in (2.1). Then*

$$f_0(x, v) = \begin{cases} (\Phi')^{-1}(E_0 - E), & E < E_0, \\ 0, & E \geq E_0, \end{cases} \quad a.e.$$

with Lagrange multiplier

$$E_0 := \frac{1}{M} \iint (E + \Phi'(f_0)) f_0 \, dv \, dx.$$

If $f_0 \in \mathcal{F}_{MC}$ is a minimizer of $\mathcal{H}$ then

$$f_0(x, v) = \begin{cases} (\Phi')^{-1}\left(\frac{E}{E_0} - 1\right), & E < E_0, \\ 0, & E \geq E_0, \end{cases} \quad a.e.$$

with Lagrange multiplier

$$E_0 := \frac{\iint E f_0 \, dv \, dx}{\iint (1 + \Phi'(f_0)) f_0 \, dv \, dx} < 0.$$

In particular, f_0 is in both cases a steady state of the Vlasov–Poisson system.

The Lagrange multiplier E_0 is negative also in case of Version 1, but the proof is different and postponed to Proposition 2.7. The choice (2.6) leads to the polytropic steady state

$$f_0(x, v) = (E_0 - E)_+^k$$

in the case of Version 1, and to a similar formula for Version 2.

PROOF OF THEOREM 2.6. We give the proof for Version 2, since due to the nonlinear nature of the constraint this case is slightly less trivial. Let f_0 and U_0 be a pointwise defined

representative of a minimizer of $\mathcal{H}$ in $\mathcal{F}_{MC}$ and of its induced potential respectively. The following abbreviation will be useful:

$$Q(f) := f + \Phi(f), \quad f \geq 0.$$

For $\varepsilon > 0$ small,

$$S_\varepsilon := \left\{ (x, v) \in \mathbb{R}^6 \mid \varepsilon \leq f_0(x, v) \leq \frac{1}{\varepsilon} \right\}$$

defines a set of positive, finite measure. Let $w \in L^\infty(\mathbb{R}^6)$ be compactly supported in $S_\varepsilon \cup f_0^{-1}(0)$ and nonnegative outside S_ε , and define

$$G(\sigma, \tau) := \iint Q(f_0 + \sigma \mathbb{1}_{S_\varepsilon} + \tau w) dv dx;$$

for τ and σ close to zero, $\tau \geq 0$, the function $f_0 + \sigma \mathbb{1}_{S_\varepsilon} + \tau w$ is bounded on S_ε , and non-negative. Therefore, G is continuously differentiable for such τ and σ , and $G(0, 0) = M$. Since

$$\partial_\sigma G(0, 0) = \iint_{S_\varepsilon} Q'(f_0) dv dx \neq 0,$$

there exists by the implicit function theorem a continuously differentiable function $\tau \mapsto \sigma(\tau)$ with $\sigma(0) = 0$, defined for $\tau \geq 0$ small, such that $G(\sigma(\tau), \tau) = M$. Hence $f_0 + \sigma(\tau) \mathbb{1}_{S_\varepsilon} + \tau w \in \mathcal{F}_{MC}$. Furthermore,

$$\sigma'(0) = -\frac{\partial_\tau G(0, 0)}{\partial_\sigma G(0, 0)} = -\frac{\iint Q'(f_0)w}{\iint_{S_\varepsilon} Q'(f_0)}. \quad (2.27)$$

Since $\mathcal{H}(f_0 + \sigma(\tau) \mathbb{1}_{S_\varepsilon} + \tau w)$ attains its minimum at $\tau = 0$,

$$0 \leq \mathcal{H}(f_0 + \sigma(\tau) \mathbb{1}_{S_\varepsilon} + \tau w) - \mathcal{H}(f_0) = \tau \iint E[\sigma'(0) \mathbb{1}_{S_\varepsilon} + w] dv dx + o(\tau)$$

for $\tau \geq 0$ small. With (2.27) we get

$$\iint [-E_\varepsilon Q'(f_0) + E]w dv dx \geq 0, \quad E_\varepsilon := \frac{\iint_{S_\varepsilon} E}{\iint_{S_\varepsilon} Q'(f_0)}.$$

By the choice for w this implies that $E = E_\varepsilon Q'(f_0)$ a.e. on S_ε and $E \geq E_\varepsilon Q'(f_0)$ a.e. on $f_0^{-1}(0)$. This shows that $E_\varepsilon = E_0$ does in fact not depend on ε . With $\varepsilon \rightarrow 0$,

$$E = E_0 Q'(f_0) \quad \text{a.e. on } f_0^{-1}(]0, \infty[),$$

$$E \geq E_0 Q'(0) = E_0 \quad \text{a.e. on } f_0^{-1}(0).$$

Multiplication of the former by f_0 and integration yields the formula for E_0 , and since

$$\iint E f_0 \, dv \, dx = E_{\text{kin}}(f_0) + 2E_{\text{pot}}(f_0) < \mathcal{H}(f_0) < 0$$

this Lagrange multiplier is negative; such a direct argument does not seem to work for Version 1 of the variational problem. $\square$

The minimizers are steady states of the Vlasov–Poisson system in the following sense: By definition, U_0 is the gravitational potential induced by f_0 . On the other hand, for a time independent potential the particle energy, and hence any function of the particle energy, is constant along characteristics and in this sense satisfies the Vlasov equation. The problem is that U_0 should be sufficiently smooth for the characteristic equations to have well-defined solutions.

PROPOSITION 2.7. *Let f_0 be a minimizer of $\mathcal{H}$ or $\mathcal{H}_C$ as obtained in Theorem 2.1 with induced spatial density ρ_0 . Alternatively, let ρ_0 be a minimizer of $\mathcal{H}_r$ as obtained in Theorem 2.4. Let U_0 be the induced potential. Then the following holds:*

(a) *the functions ρ_0 and U_0 are spherically symmetric with respect to some point in $\mathbb{R}^3$, and ρ_0 is decreasing as a function of the radial variable;*

(b) *$\rho_0 \in C_c(\mathbb{R}^3)$, $U_0 \in C^2(\mathbb{R}^3)$ with $\lim_{|x| \rightarrow \infty} U_0(x) = 0$, and $E_0 < 0$. If ρ_0 comes from a minimizer of $\mathcal{H}_C$ or $\mathcal{H}$ then $\rho_0 \in C_c^1(\mathbb{R}^3)$. Minimizers f_0 are compactly supported also with respect to v .*

PROOF. In order to prove part (a) we consider first the case that $\rho_0 \in \mathcal{R}_M$ is a minimizer of the reduced functional $\mathcal{H}_r$ as obtained in Theorem 2.4. As observed in Step 5 of the proof of that theorem ρ_0 is spherically symmetric with respect to some point in $\mathbb{R}^3$, and decreasing as a function of the radial variable. Let f_0 be a minimizer of the energy-Casimir functional $\mathcal{H}_C$. Then the assertions for ρ_0 and U_0 remain true, since f_0 arises from a minimizer of the reduced functional by the lifting process in Theorem 2.2(b).

To prove the spherical symmetry of a minimizer f_0 of the energy $\mathcal{H}$ we denote by f_0^* its spherically symmetric rearrangement with respect to x . Arguing as above, $f_0(x, v) = f_0^*(x + a_v, v)$ for some possibly v -dependent shift vector a_v . Since both f_0 and f_0^* are minimizers they are both of the form stated in Theorem 2.6, so $E_0 \Phi'(f_0(x, v)) = \frac{1}{2}|v|^2 + U_{f_0}(x) - E_0$ and $E_0^* \Phi'(f_0^*(x, v)) = \frac{1}{2}|v|^2 + U_{f_0^*}(x) - E_0^*$. The explicit form of E_0 implies that $E_0 = E_0^*$, hence $U_{f_0}(x) = U_{f_0^*}(x + a_v)$, and a_v is independent of v . Hence the minimizer f_0 is a spatial translation of f_0^* , which proves the symmetry assertion.

As to part (b), we note first that by Theorem 2.6 a minimizer f_0 obtained in Theorem 2.1 satisfies a relation of the form

$$f_0(x, v) = \phi(E(x, v))$$

with ϕ determined by the function Φ and E_0 . This in turn implies a relation between ρ_0 and U_0 ,

$$\rho_0(x) = h(U_0(x)) := 4\pi\sqrt{2} \int_{U_0(x)}^{\infty} \phi(E)\sqrt{E - U_0(x)} \, dE, \quad (2.28)$$

where h is continuously differentiable. For a minimizer ρ_0 of $\mathcal{H}_r$ such a relation holds by (2.12); in this case h is determined by Ψ and need only be continuous.

Let $\rho_0 \in L_+^p(\mathbb{R}^3)$ for some $p > 1$, and as usual $1/p + 1/q = 1$. For any $R > 1$ we split the convolution integral defining U_0 according to $|x - y| < 1/R$, $1/R \leq |x - y| < R$, and $|x - y| \geq R$ to obtain

$$-U_0(x) \leq C\|\rho_0\|_p \left(\int_0^{1/R} r^{2-q} \, dr \right)^{1/q} + R \int_{|y| \geq |x| - R} \rho_0(y) \, dy + \frac{M}{R}.$$

This implies that $U_0 \in L^\infty(\mathbb{R}^3)$ with $U_0(x) \rightarrow 0$, $|x| \rightarrow \infty$, provided $q < 3$, i.e., $p > 3/2$. Assume for the moment that this is true. Then by (2.28), $\rho_0 \in L^1 \cap L^\infty(\mathbb{R}^3)$. By spherical symmetry,

$$U_0(r) = -\frac{4\pi}{r} \int_0^r s^2 \rho_0(s) \, ds - 4\pi \int_r^\infty s \rho_0(s) \, ds, \quad U_0'(r) = \frac{4\pi}{r^2} \int_0^r s^2 \rho_0(s) \, ds,$$

where $r = |x|$, in particular U_0 is continuous. Again by (2.28), ρ_0 is continuous as well, and the formulas above imply the asserted regularity of U_0 .

If $0 < n < 2$ then $p := 1 + 1/n > 3/2$, and the regularity assumptions are established. If $2 \leq n < 5$ a little more work is required. By the assumptions $(\Phi 1)$ and $(\Phi 2)$ and the mean value theorem,

$$\Phi'(f) \geq \Phi'(\tau) = \frac{\Phi(f) - \Phi(0)}{f - 0} \geq Cf^{1/k}$$

for all f large, with some intermediate value $0 \leq \tau \leq f$. Similarly, for ρ large,

$$\Psi'(\rho) \geq C\rho^{1/n}.$$

In both cases the relation (2.28) together with these estimates imply that

$$\rho_0(x) \leq C(1 + (E_0 - U_0(x))_+^n).$$

If we use this estimate on the set of finite measure where ρ_0 is large and the integrability of ρ_0 on the complement we find that

$$\int \rho_0(x)^p \, dx \leq C + C \int (-U_0(x))^{np} \, dx. \quad (2.29)$$

Starting with $p_0 = 1 + 1/n$ we apply Lemma P2(a) to find that U_0 lies in L^q with $q = (1/p_0 - 2/3)^{-1} > 1$, and substituting this into (2.29) we conclude that $\rho_0 \in L^{p_1}$ with $p_1 = q/n$; note that by assumption $p_0 < 3/2$. If $p_1 > 3/2$ we are done. If $p_1 = 3/2$ we decrease p_1 slightly – note that $\rho_0 \in L^1$ – so that in the next step we find p_2 as large as we wish. If $p_1 < 3/2$ we repeat the process. By induction,

$$p_k = \frac{3(1 + 1/n)(n - 1)}{n^k(n - 5) + 2n + 2} > 1$$

as long as $p_{k-1} < 3/2$. But since $2 \leq n < 5$ the denominator would eventually become negative so that the process must stop after finitely many steps, and again $\rho_0 \in L^p(\mathbb{R}^3)$ for some $p > 3/2$.

The minimizer of $\mathcal{H}_C$ or $\mathcal{H}_r$ obtained in Theorem 2.1 or Theorem 2.4 has compact support by Step 3 of the proof of the latter theorem. The limiting behavior of U_0 together with Theorem 2.6 implies that $E_0 < 0$.

For Version 2 of the variational problem $E_0 < 0$ by Theorem 2.6. Hence $\lim_{|x| \rightarrow \infty} U_0(x) = 0$ implies that for $|x|$ sufficiently large, $E(x, v) > E_0$, and by Theorem 2.6, f_0 and ρ_0 have compact support also in this case. $\square$

A question which is of interest in itself and which is also relevant for the stability discussion in the next section is the possible uniqueness or nonuniqueness of the minimizer. So far, only preliminary results in this direction exist.

REMARK. (a) Consider the polytropic case $\Phi(f) = f^{1+1/k}$ or $\Psi(\rho) = \rho^{1+1/n}$, respectively. If $0 < k < 3/2$ or $0 < n < 3$ then up to spatial translations the functional $\mathcal{H}_C$ or $\mathcal{H}_r$ has exactly one minimizer with prescribed mass $M > 0$. If $0 < k < 7/2$ then up to spatial translations the energy $\mathcal{H}$ has at most two minimizers in the constraint set $\mathcal{F}_{MC}$.

We show the uniqueness assertion under the mass constraint; the proof under the mass-Casimir constraint is more technical, cf. [42], Theorem 3. Up to some shift U_0 as a function of the radial variable $r := |x|$ solves the equation

$$\frac{1}{r^2} (r^2 U_0')' = c(E_0 - U_0)_+^n, \quad r > 0, \quad (2.30)$$

with some appropriately defined constant $c > 0$. The function $E_0 - U_0$ is a solution of the singular ordinary differential equation

$$\frac{1}{r^2} (r^2 z')' = -cz_+^n, \quad r > 0. \quad (2.31)$$

Solutions $z \in C([0, \infty]) \cap C^2(]0, \infty[)$ of (2.31) with z' bounded near $r = 0$ are uniquely determined by $z(0)$. If z is such a solution then so is

$$z_\alpha(r) := \alpha z(\alpha^\gamma r), \quad r \geq 0,$$

for any $\alpha > 0$ where $\gamma := (n - 1)/2$, and $z_\alpha(0) = \alpha z(0)$. Now assume there exists another minimizer with mass M , i.e., up to a shift another solution U_1 of (2.30) with cut-off energy E_1 . Uniqueness for (2.31) yields some $\alpha > 0$ such that

$$E_1 - U_1(r) = \alpha E_0 - \alpha U_0(\alpha^\gamma r), \quad r \geq 0.$$

However, both steady states have the same mass M , so that

$$\begin{aligned} M &= c \int_0^\infty r^2 (E_1 - U_1(r))_+^n dr \\ &= \alpha^{n-3\gamma} c \int_0^\infty r^2 (E_0 - U_0(r))_+^n dr = \alpha^{n-3\gamma} M. \end{aligned}$$

For $0 < n < 3$ the exponent of α is not zero, hence $\alpha = 1$, and considering limits at spatial infinity we conclude that $E_0 = E_1$ and $U_0 = U_1$.

(b) Let Ψ be such that $\Psi(0) = 0$ and

$$\Psi'(\rho) = \begin{cases} \rho, & 0 \leq \rho \leq 1, \\ \rho^{1/10}, & 1 < \rho < 10, \\ 10^{-9/10} \rho, & 10 \leq \rho. \end{cases}$$

This function satisfies $(\Psi 1)$, $(\Psi 2)$, $(\Psi 3)$; note however that the exponent used for $1 < \rho < 10$ corresponds to $n = 10$ which is well outside of the required range $0 < n < 3$. If the resulting equation for $z = E_0 - U_0$ is solved numerically then the choices $z(0) = 0.522, 1.641, 2.364$ give three different steady states with the same mass $M = 0.462$. The minimizers of $\mathcal{H}_r$ for this value of M must be among these three states, and it turns out that the values of $\mathcal{H}_r$ resulting from $z(0) = 0.522, 2.364$ are equal and smaller than the one resulting from $z(0) = 1.641$. Hence for this example there are two distinct minimizers. Clearly, this also provides a counterexample to uniqueness of the minimizer for the energy-Casimir functional $\mathcal{H}_C$.

A similar example of nonuniqueness of the minimizer of $\mathcal{H}$ is reported in [107], Section 5. We have found no numerical indication that under our general assumptions there might be infinitely many minimizers. In particular, minimizers always seem to be isolated.

2.6. Dynamical stability

We now come to the stability assertion for the steady states which are obtained as minimizers above. To this end we first rewrite the Taylor expansion of the energy or energy-Casimir functional, respectively.

REMARK. (a) In case of Version 1,

$$\mathcal{H}_C(f) - \mathcal{H}_C(f_0) = d(f, f_0) - \frac{1}{8\pi} \int |\nabla U_f - \nabla U_0|^2 dx, \quad (2.32)$$

where for $f \in \mathcal{F}_M$,

$$\begin{aligned} d(f, f_0) &:= \iint [\Phi(f) - \Phi(f_0) + E(f - f_0)] \, dv \, dx \\ &= \iint [\Phi(f) - \Phi(f_0) + (E - E_0)(f - f_0)] \, dv \, dx \\ &\geq \iint [\Phi'(f_0) + (E - E_0)](f - f_0) \, dv \, dx \geq 0 \end{aligned}$$

with $d(f, f_0) = 0$ iff $f = f_0$.

(b) In case of Version 2,

$$\mathcal{H}(f) - \mathcal{H}(f_0) = d(f, f_0) - \frac{1}{8\pi} \int |\nabla U_f - \nabla U_0|^2 \, dx, \quad (2.33)$$

where for $f \in \mathcal{F}_{MC}$,

$$\begin{aligned} d(f, f_0) &:= \iint E(f - f_0) \, dv \, dx \\ &= \iint [(-E_0)(\Phi(f) - \Phi(f_0)) + (E - E_0)(f - f_0)] \, dv \, dx \\ &\geq \iint [(-E_0)\Phi'(f_0) + (E - E_0)](f - f_0) \, dv \, dx \geq 0 \end{aligned}$$

with $d(f, f_0) = 0$ iff $f = f_0$.

This is due to the strict convexity of Φ , and the fact that on the support of f_0 the bracket vanishes by Theorem 2.6; note also that in the second equality we added a zero due to the respective constraint.

THEOREM 2.8. *Let f_0 be a minimizer as obtained in Theorem 2.1, in case of Version 1 assume that the minimizer is unique or at least isolated up to shifts in x . Then the following nonlinear stability assertion holds:*

For any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any classical solution $t \mapsto f(t)$ of the Vlasov–Poisson system with $f(0) \in C_c^1(\mathbb{R}^6) \cap \mathcal{F}_M$ or $f(0) \in C_c^1(\mathbb{R}^6) \cap \mathcal{F}_{MC}$, respectively, the initial estimate

$$d(f(0), f_0) + \frac{1}{8\pi} \int |\nabla U_{f(0)} - \nabla U_0|^2 \, dx < \delta$$

implies that for any $t \geq 0$ there is a shift vector $a \in \mathbb{R}^3$ such that

$$d(T^a f(t), f_0) + \frac{1}{8\pi} \int |T^a \nabla U_{f(t)} - \nabla U_0|^2 \, dx < \varepsilon, \quad t \geq 0.$$

As above, $T^a f(x, v) := f(x + a, v)$.

PROOF. Let us first assume that the minimizer is unique up to spatial translations, and let us consider Version 1 first. Assume the assertion is false. Then there exist $\varepsilon > 0$, $t_j > 0$, $f_j(0) \in C_c^1(\mathbb{R}^6) \cap \mathcal{F}_M$ such that for $j \in \mathbb{N}$,

$$d(f_j(0), f_0) + \frac{1}{8\pi} \int |\nabla U_{f_j(0)} - \nabla U_0|^2 dx < \frac{1}{j},$$

but for any shift vector $a \in \mathbb{R}^3$,

$$d(T^a f_j(t_j), f_0) + \frac{1}{8\pi} \int |T^a \nabla U_{f_j(t_j)} - \nabla U_0|^2 dx \geq \varepsilon.$$

Since $\mathcal{H}_C$ is conserved, (2.32) and the assumption on the initial data imply that $\mathcal{H}_C(f_j(t_j)) = \mathcal{H}_C(f_j(0)) \rightarrow \mathcal{H}_C(f_0)$, i.e., $(f_j(t_j)) \subset \mathcal{F}_M$ is a minimizing sequence. Hence by Theorem 2.1, $\int |\nabla U_{f_j(t_j)} - \nabla U_0|^2 \rightarrow 0$ up to subsequences and shifts in x , provided that there is no other minimizer to which this sequence can converge. By (2.32), $d(f_j(t_j), f_0) \rightarrow 0$ as well, which is the desired contradiction. If the minimizer is unique up to shifts, the proof for Version 2 is completely analogous.

By definition, we call the minimizer isolated up to spatial translations if

$$\inf\{\|\nabla U_{f_0} - \nabla U_{\tilde{f}_0}\|_2 \mid \tilde{f}_0 \in \mathcal{M}_M \setminus \{T^a f_0 \mid a \in \mathbb{R}^3\}\} > 0,$$

where $\mathcal{M}_M$ denotes the set of all minimizers of the given functional under the given constraint. The argument above then has to be combined with a continuity argument to show that the assertion of the theorem still holds true, cf. [98], p. 124. For Version 2 a much less trivial argument due to Schaeffer [107] shows that the theorem remains true even if the minimizer is not isolated. $\square$

The spatial shifts appearing in the stability statement are again due to the spatial invariance of the system. If f_0 is perturbed by giving all the particles an additional, fixed velocity, then in space the corresponding solution travels off from f_0 at a linear rate in t , no matter how small the perturbation. Hence without the spatial shifts the assertion of the theorem is false. A stability result of this type is sometimes referred to as *orbital stability*, cf. [72, 104].

A weak point of the present approach is the fact that the proof is not constructive – given ε it is not known how small the corresponding δ must be. A nice feature of the result is that the same quantity is used to measure the deviation initially and at later times t . In infinite-dimensional dynamical systems initial control in a strong norm can be necessary to gain control in a weaker norm at later times. On the other hand, it certainly is desirable to achieve the stability estimate also in some norm for f . In [72, 104] results in this direction are obtained by changing the variational approach. However, such improvements are easily obtained within the framework presented here. To see this we have to think for a moment about what perturbations are admissible from a physics point of view.

Remark on dynamically accessible perturbations. A galaxy in equilibrium represented by a steady state f_0 is typically perturbed by the gravitational pull of some (distant) outside object like a neighboring galaxy. This means that an external force acts on the particles in addition to the self-consistent one. The resulting perturbation simply consists in a reshuffling of the particles in phase space. Hence a physically natural class of perturbations are all states f which are equimeasurable to f_0 , $f \sim f_0$, which by definition means that

$$\forall \tau \geq 0: \quad \text{vol}(\{(x, v) \in \mathbb{R}^6 \mid f(x, v) > \tau\}) = \text{vol}(\{(x, v) \in \mathbb{R}^6 \mid f_0(x, v) > \tau\}).$$

Notice that this class is invariant under the Vlasov–Poisson system. Clearly, if $f \sim f_0$ then $\|f\|_p = \|f_0\|_p$ for any $p \in [1, \infty]$.

With this remark in mind we arrive at the following stronger stability result; notice that we need not even exploit the full strength of the restriction on the perturbations introduced above.

COROLLARY 2.9. *If in Theorem 2.8 the assumption $\|f(0)\|_{1+1/k} = \|f_0\|_{1+1/k}$ is added then for any $\varepsilon > 0$ the parameter $\delta > 0$ can be chosen such that the additional stability estimate*

$$\|T^a f(t) - f_0\|_{1+1/k} < \varepsilon, \quad t \geq 0,$$

holds. If $K > 0$ is such that $\text{vol}(\text{supp } f_0) < K$, then this stability estimate holds with any $p \in [1, 1 + 1/k]$ instead of $1 + 1/k$, provided the perturbations satisfy the additional restriction $\text{vol}(\text{supp } f(0)) < K$. If $K > 0$ is such that $\|f_0\|_\infty < K$ and the perturbations satisfy the restriction $\|f(0)\|_\infty < K$ then the same is true for any $p \in [1 + 1/k, \infty[$.

PROOF. We repeat the proof of Theorem 2.8 except that in the contradiction assumption we have

$$\begin{aligned} & \|T^a f_j(t_j) - f_0\|_{1+1/k} + d(T^a f_j(t_j), f_0) + \frac{1}{8\pi} \int |T^a \nabla U_{f_j(t_j)} - \nabla U_0|^2 dx \\ & \geq \varepsilon. \end{aligned}$$

Now we observe that from the minimizing sequence $(f_j(t_j))$ obtained in that proof we can extract a subsequence which converges weakly in $L^{1+1/k}$ to f_0 by Theorem 2.1. But due to our additional restriction on the perturbations

$$\|f_j(t_j)\|_{1+1/k} = \|f_0\|_{1+1/k}, \quad j \in \mathbb{N}.$$

By the Radon–Riesz–Theorem [74], Theorem 2.11, this implies that $f_j(t_j) \rightarrow f_0$ strongly in $L^{1+1/k}$. Together with the rest of the proof of Theorem 2.8 this proves the first assertion. Under the additional restriction on the perturbations, $\text{vol}(\text{supp } f(t)) < K$ or $\|f(t)\|_\infty < K$ for all times, and the additional assertions follow by Hölder’s inequality and interpolation. $\square$

Concluding remarks. (a) The conditions for stability are formulated in terms of the Casimir function Φ , but they can be translated into conditions on the steady state $f_0 = \phi(E)$. In particular, the crucial assumption that Φ be strictly convex means that ϕ is strictly decreasing on its support.

(b) In [47] spherically symmetric steady states depending also on the modulus of angular momentum L , defined in (2.3), are dealt with. In [49] axially symmetric minimizers depending on the particle angular momentum corresponding to the axis of symmetry are considered. The method yields stability against perturbations which respect the symmetry, but not against general perturbations. This is due to the fact that the function Φ in the Casimir functional $\mathcal{C}$ must in these cases depend on the additional particle invariant, and hence the Casimir functional is preserved only along solutions with the proper symmetry. Stability of nonisotropic steady states which for example depend also on L against nonsymmetric perturbations is an interesting open problem, in particular, since in view of the above discussion of dynamically accessible perturbations symmetry restrictions are unphysical and at best are mathematical stepping stones toward more satisfactory results.

(c) A similar problem arises with flat steady states where all the particles are restricted to a plane. They are used as models for extremely flattened, disk-like galaxies. Their stability was investigated by variational techniques in [94], but the perturbations were restricted to live in the plane.

(d) So far, no rigorous instability results are known in the stellar dynamics case for steady states which violate the stability conditions. Such results do exist in the plasma physics case, cf. [51, 52, 75–78].

(e) By similar techniques a preliminary result toward stability was established for the Vlasov–Einstein system [113].

(f) The above stability result brings up a question concerning the initial value problem for the Vlasov–Poisson system: Can one extend the class of admissible initial data in such a way that it contains the steady states considered above? Notice that f_0 need not be continuously differentiable. Going further one might wish to admit all dynamically accessible perturbations originating from these steady states as initial data. It is not hard to establish stability results within the context of weak solutions in the sense that the stability estimates then hold for such weak solutions which are obtained as limits of solutions to certain regularized systems, cf. [67, 72]. However, due to the inherent nonuniqueness such a formulation is unsatisfactory. It is therefore desirable to have a global existence and uniqueness result which covers these states as initial data and provides solutions which preserve all the conserved quantities, cf. [114].

(g) An ansatz of the form

$$f_0(x, v) = (e^{E_0 - E} - 1)_+$$

also leads to a steady state with compact support and finite mass, cf. [102]. This so-called King model is important in astrophysics, but since the corresponding Casimir function

$$\Phi(f) = (1 + f) \ln(1 + f) - f$$

does not satisfy the growth condition $(\Phi 2)$, it cannot be dealt with by the above variational approach. A nonvariational approach which covers the King model has recently been developed in [50].

2.7. The reduced variational problem and the Euler–Poisson system

So far the reduced variational problem played the role of a mathematical device. In the present section we demonstrate that the reduction procedure is much more than that: It points to a deep connection between the Vlasov–Poisson and the Euler–Poisson systems on the level of their steady states and their stability.

If $\rho_0 \in \mathcal{R}_M$ minimizes the reduced functional $\mathcal{H}_r$, then ρ_0 supplemented with the velocity field $u_0 = 0$ is a steady state of the Euler–Poisson system

$$\begin{aligned}\partial_t \rho + \operatorname{div}(\rho u) &= 0, \\ \rho \partial_t u + \rho(u \cdot \partial_x)u &= -\partial_x p - \rho \partial_x U, \\ \Delta U &= 4\pi \rho, \quad \lim_{|x| \rightarrow \infty} U(t, x) = 0,\end{aligned}$$

with equation of state

$$p = P(\rho) := \rho \Psi'(\rho) - \Psi(\rho).$$

This follows from the Euler–Lagrange identity (2.12). Here u and p denote the velocity field and the pressure of an ideal, compressible fluid with mass density ρ , and the fluid self-interacts via its induced gravitational potential U . This system is sometimes used as a simple model for a gaseous, barotropic star. The beautiful thing now is that the state $(\rho_0, u_0 = 0)$ obviously minimizes the energy

$$\mathcal{H}(\rho, u) := \frac{1}{2} \int |u|^2 \rho \, dx + \int \Psi(\rho) \, dx + E_{\text{pot}}(\rho)$$

of the system, which is a conserved quantity. Expanding as before we find that

$$\mathcal{H}(\rho, u) - \mathcal{H}(\rho_0, 0) = \frac{1}{2} \int |u|^2 \rho \, dx + d(\rho, \rho_0) - \frac{1}{8\pi} \int |\nabla U_\rho - \nabla U_0|^2 \, dx,$$

where for $\rho \in \mathcal{R}_M$,

$$d(\rho, \rho_0) := \int [\Psi(\rho) - \Psi(\rho_0) + (U_0 - E_0)(\rho - \rho_0)] \, dx \geq 0,$$

with equality iff $\rho = \rho_0$. The same proof as for the Vlasov–Poisson system implies a stability result for the Euler–Poisson system – the term with the unfavorable sign in the expansion again tends to zero along minimizing sequences, cf. Theorem 2.4. However, there is an

important caveat: While for the Vlasov–Poisson system we have global-in-time solutions for sufficiently nice data, and these solutions really preserve all the conserved quantities, no such result is available for the Euler–Poisson system, and we only obtain a

CONDITIONAL STABILITY RESULT. For every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every solution $t \mapsto (\rho(t), u(t))$ with $\rho(0) \in \mathcal{R}_M$ which preserves energy and mass the initial estimate

$$\frac{1}{2} \int |u(0)|^2 \rho(0) \, dx + d(\rho(0), \rho_0) + \frac{1}{8\pi} \int |\nabla U_{\rho(0)} - \nabla U_0|^2 \, dx < \delta$$

implies that as long as the solution exists,

$$\frac{1}{2} \int |u(t)|^2 \rho(t) \, dx + d(\rho(t), \rho_0) + \frac{1}{8\pi} \int |\nabla U_{\rho(t)} - \nabla U_0|^2 \, dx < \varepsilon$$

up to shifts in x and provided the minimizer is unique up to such shifts.

The same comments as on Theorem 2.8 apply. Because of the above caveat we prefer not to call this a theorem, although as far as the stability analysis itself is concerned it is perfectly rigorous. The open problem is whether a suitable concept of solution to the initial value problem exists.

The relation between the fluid and the kinetic steady states. Now that minimizers of the reduced functional are identified as stable steady states of the Euler–Poisson system it is instructive to reconsider the reduction procedure leading from the kinetic to the fluid dynamics picture. First we recall that for the Legendre transform $\bar{h}$ of a function h the following holds:

$$h'(\xi) = \eta \iff h(\xi) + \bar{h}(\eta) = \xi\eta \iff (\bar{h})'(\eta) = \xi.$$

If f_0 is a minimizer of $\mathcal{H}_C$,

$$\begin{aligned} f_0 &= (\Phi')^{-1}(E_0 - E) = (\bar{\Phi})'(E_0 - E), \\ \rho_0 &= \int f_0 \, dv = \int (\bar{\Phi})' \left(E_0 - U_0 - \frac{1}{2}|v|^2 \right) \, dv, \end{aligned}$$

and

$$p_0 = \frac{1}{3} \int |v|^2 f_0 \, dv = \int \bar{\Phi} \left(E_0 - U_0 - \frac{1}{2}|v|^2 \right) \, dv$$

is the induced, isotropic pressure. On the other hand, if ρ_0 is a minimizer of the reduced functional $\mathcal{H}_r$,

$$\begin{aligned} \rho_0 &= (\Psi')^{-1}(E_0 - U_0) = (\bar{\Psi})'(E_0 - U_0), \\ p_0 &= P(\rho_0) = \rho_0 \Psi'(\rho_0) - \Psi(\rho_0) = \bar{\Psi}(\Psi'(\rho_0)) = \bar{\Psi}(E_0 - U_0). \end{aligned}$$

In both the kinetic and the fluid picture the spatial density and the pressure are functionals of the potential, and these functional relations on the kinetic and on the fluid level fit provided

$$\bar{\Psi}(\lambda) = \int \bar{\Phi}\left(\lambda - \frac{1}{2}|v|^2\right) dv,$$

which is exactly the relation between Φ and Ψ obtained by the reduction mechanism.

The threshold $k = 3/2$, $n = 3$. It is worthwhile to review the role of the threshold $k = 3/2$, $n = 3$ in the context of the relation between the Vlasov–Poisson and the Euler–Poisson system. For the Vlasov–Poisson system the Casimir functional $\mathcal{C}$ is preserved, and hence it is possible to incorporate $\mathcal{C}$ into the functional to be minimized or into the constraint. We have seen that the former approach, which allows for reduction, works only for $0 < k < 3/2$ while the latter works for $0 < k < 7/2$. That this is not just a mathematical technicality can be seen from the following observation.

By Theorem 2.2 the energy–Casimir functional in the kinetic picture equals the energy functional in the fluid picture in the case of a minimizer. For polytropes, $\text{sign } \mathcal{H}_C(f_0) = \text{sign}(n - 3)$, i.e., the energy–Casimir functional in the kinetic and the energy in the fluid picture changes sign at $n = 3$. The energy $\mathcal{H}$ in the kinetic picture however remains negative for $0 < k < 7/2$. Secondly, if the perturbation of a steady state has positive energy then this perturbation is unstable in the sense of Corollary 1.16 from Chapter 1. An analogous result holds for the Euler–Poisson system, cf. [19]. Hence stability is lost for the Euler–Poisson system at $n = 3$ and so reduction in the sense we used it cannot work for $k \geq 3/2$.

A nonlinear instability result for the Euler–Poisson system with equation of state $p = A\rho^{6/5}$ was recently established in [64]. Notice that this equation of state corresponds to $n = 5$, which is well outside the range of stability which was established above for the Euler–Poisson case. On the other hand, the corresponding Vlasov–Poisson steady state $f_0(x, v) = c(-E)_+^{7/2}$, the so-called Plummer sphere, is stable, cf. [48], Section 6. That this state is the minimizer of the energy under an appropriate constraint had been observed earlier in [1], cf. also [2].

References

- [1] J.J. Aly, *On the lowest energy state of a collisionless selfgravitating system under phase space volume constraints*, Mon. Not. R. Astron. Soc. **241** (1989), 15–27.
- [2] J.J. Aly, *Existence of a minimum energy state for a constrained collisionless gravitational system*, Preprint (2005).
- [3] H. Andréasson, *The Einstein–Vlasov system/kinetic theory*, Living Reviews in Relativity, Max Planck Institute for Gravitational Physics (2005); available at <http://www.livingreviews.org/lrr-2005-2>.
- [4] A.A. Arsen'ev, *Global existence of a weak solution of Vlasov's system of equations*. Comput. Math. Math. Phys. **15** (1975), 131–141.
- [5] C. Bardos and P. Degond, *Global existence for the Vlasov–Poisson system in 3 space variables with small initial data*, Ann. Inst. H. Poincaré Anal. Non Linéaire **2** (1985), 101–118.
- [6] J. Batt, *Ein Existenzbeweis für die Vlasov–Gleichung der Stelldynamik bei gemittelter Dichte*, Arch. Ration. Mech. Anal. **13** (1963), 296–308.

- [7] J. Batt, *Global symmetric solutions of the initial value problem in stellar dynamics*, J. Differential Equations **25** (1977), 342–364.
- [8] J. Batt, H. Berestycki, P. Degond and B. Perthame, *Some families of solutions of the Vlasov–Poisson system*, Arch. Ration. Mech. Anal. **104** (1988), 79–103.
- [9] J. Batt, W. Faltenbacher and E. Horst, *Stationary spherically symmetric models in stellar dynamics*, Arch. Ration. Mech. Anal. **93** (1986), 159–183.
- [10] J. Batt, M. Kunze and G. Rein, *On the asymptotic behaviour of a one-dimensional, monocharged plasma*, Adv. Differential Equations **3** (1998), 271–292.
- [11] J. Batt, P. Morrison and G. Rein, *Linear stability of stationary solutions of the Vlasov–Poisson system in three dimensions*, Arch. Ration. Mech. Anal. **130** (1995), 163–182.
- [12] J. Batt and G. Rein, *Global classical solutions of the periodic Vlasov–Poisson system in three dimensions*, C. R. Acad. Sci. Paris **313** (1991), 411–416.
- [13] J. Binney and S. Tremaine, *Galactic Dynamics*, Princeton Univ. Press, Princeton, NJ (1987).
- [14] F. Bouchut, *Existence and uniqueness of a global smooth solution for the Vlasov–Poisson–Fokker–Planck system in three dimensions*, J. Funct. Anal. **111** (1993), 239–258.
- [15] F. Bouchut, F. Golse and C. Pallard, *Classical solutions and the Glassey–Strauss theorem for the 3d Vlasov–Maxwell system*, Arch. Ration. Mech. Anal. **170** (2003), 1–15.
- [16] A. Burchard and Y. Guo, *Compactness via symmetrization*, J. Funct. Anal. **214** (2004), 40–73.
- [17] S. Calogero, *Global classical solutions to the 3D Nordström–Vlasov system*, Preprint (2005); available at [arXiv:math-ph/0507030](https://arxiv.org/abs/math-ph/0507030).
- [18] C. Cercignani, R. Illner and M. Pulvirenti, *The Mathematical Theory of Dilute Gases*, Springer-Verlag, New York (1994).
- [19] Y. Deng, T.-P. Liu, T. Yang and Z.-A. Yao, *Solutions of Euler–Poisson equations for gaseous stars*, Arch. Ration. Mech. Anal. **164** (2002), 261–285.
- [20] C. Dietz and V. Sandor, *The hydrodynamical limit of the Vlasov–Poisson system*, Transport Theory Statist. Phys. **28** (1999), 499–520.
- [21] R.J. DiPerna and P.-L. Lions, *Global weak solutions of Vlasov–Maxwell systems*, Comm. Pure Appl. Math. **XLII** (1989), 729–757.
- [22] R.J. DiPerna and P.-L. Lions, *On the Cauchy problem for Boltzmann equations: Global existence and weak stability*, Ann. of Math. **130** (1989), 321–366.
- [23] J. Dolbeault and G. Rein, *Time-dependent rescalings and Lyapunov functionals for the Vlasov–Poisson and Euler–Poisson systems, and for related models of kinetic equations, fluid dynamics and quantum physics*, Math. Models Methods Appl. Sci. **11** (2001), 407–432.
- [24] J. Dolbeault, O. Sanchez and J. Soler, *Asymptotic behaviour for the Vlasov–Poisson system in the stellar-dynamics case*, Arch. Ration. Mech. Anal. **171** (2004), 301–327.
- [25] J. Ehlers and W. Rienstra, *The locally isotropic solutions of the Liouville and Poisson equations*, Astrophys. J. **155** (1969), 105–116.
- [26] A.M. Fridman and V.L. Polyachenko, *Physics of Gravitating Systems I*, Springer-Verlag, New York (1984).
- [27] I. Gasser, P.-E. Jabin and B. Perthame, *Regularity and propagation of moments in some nonlinear Vlasov systems*, Proc. Roy. Soc. Edinburgh Sect. A **130** (2000), 1259–1273.
- [28] R.T. Glassey, *The Cauchy Problem in Kinetic Theory*, SIAM, Philadelphia, PA (1996).
- [29] R.T. Glassey and J. Schaeffer, *On symmetric solutions of the relativistic Vlasov–Poisson system*, Comm. Math. Phys. **101** (1985), 459–473.
- [30] R.T. Glassey and J. Schaeffer, *Global existence of the relativistic Vlasov–Maxwell system with nearly neutral initial data*, Comm. Math. Phys. **119** (1988), 353–384.
- [31] R.T. Glassey and J. Schaeffer, *On the “one and one-half dimensional” relativistic Vlasov–Maxwell system*, Math. Methods Appl. Sci. **13** (1990), 169–179.
- [32] R.T. Glassey and J. Schaeffer, *The “two and one-half dimensional” relativistic Vlasov–Maxwell system*, Comm. Math. Phys. **185** (1997), 257–284.
- [33] R.T. Glassey and J. Schaeffer, *The relativistic Vlasov–Maxwell system in two space dimensions, I and II*, Arch. Ration. Mech. Anal. **141** (1998), 331–354; 355–374.
- [34] R.T. Glassey and J. Schaeffer, *On global symmetric solutions of the relativistic Vlasov–Poisson equation in three space dimensions*, Math. Methods Appl. Sci. **24** (2001), 143–157.
- [35] R.T. Glassey and W.A. Strauss, *Remarks on collisionless plasmas*, Contemp. Math. **28** (1984), 269–279.

- [36] R.T. Glassey and W.A. Strauss, *Singularity formation in a collisionless plasma could occur only at high velocities*, Arch. Ration. Mech. Anal. **92** (1986), 59–90.
- [37] R.T. Glassey and W.A. Strauss, *Absence of shocks in an initially dilute collisionless plasma*, Comm. Math. Phys. **113** (1987), 191–208.
- [38] Y. Guo, *Global weak solutions to the Vlasov–Maxwell system with boundary conditions*, Comm. Math. Phys. **154** (1993), 245–263.
- [39] Y. Guo, *Regularity for the Vlasov equations in a half space*, Indiana Univ. Math. J. **43** (1994), 255–320.
- [40] Y. Guo, *Singular solutions of the Vlasov–Maxwell system on a half line*, Arch. Ration. Mech. Anal. **131** (1995), 241–304.
- [41] Y. Guo, *Variational method in polytropic galaxies*, Arch. Ration. Mech. Anal. **150** (1999), 209–224.
- [42] Y. Guo, *On the generalized Antonov’s stability criterion*, Contemp. Math. **263** (2000), 85–107.
- [43] Y. Guo, *The Vlasov–Poisson–Boltzmann system near vacuum*, Comm. Math. Phys. **218** (2001), 293–313.
- [44] Y. Guo, *The Vlasov–Maxwell–Boltzmann system near Maxwellians*, Invent. Math. **153** (2003), 593–630.
- [45] Y. Guo and C. Grotta Ragazzo, *On steady states in a collisionless plasma*, Comm. Pure Appl. Math. **XVIX** (1996), 1145–1174.
- [46] Y. Guo and G. Rein, *Stable steady states in stellar dynamics*, Arch. Ration. Mech. Anal. **147** (1999), 225–243.
- [47] Y. Guo and G. Rein, *Existence and stability of Camm type steady states in galactic dynamics*, Indiana Univ. Math. J. **48** (1999), 1237–1255.
- [48] Y. Guo and G. Rein, *Isotropic steady states in galactic dynamics*, Comm. Math. Phys. **219** (2001), 607–629.
- [49] Y. Guo and G. Rein, *Stable models of elliptical galaxies*, Mon. Not. R. Astron. Soc. **344** (2003), 1396–1406.
- [50] Y. Guo and G. Rein, *A non-variational approach to nonlinear stability in stellar dynamics applied to the King model*, Preprint (2006); available at [arXiv:math-ph/0602058](https://arxiv.org/abs/math-ph/0602058).
- [51] Y. Guo and W.A. Strauss, *Nonlinear instability of double-humped equilibria*, Ann. Inst. H. Poincaré Anal. Non Linéaire **12** (1995), 339–352.
- [52] Y. Guo and W.A. Strauss, *Unstable BGK solitary waves and collisionless shocks*, Comm. Math. Phys. **195** (1998), 267–293.
- [53] M. Hadžić, *Compactness and stability of some systems of nonlinear PDE-s in galactic dynamics*, Diploma thesis, University of Vienna (2005).
- [54] M. Hauray and P.-E. Jabin, *N-particle approximation of the Vlasov equations with singular potential*, Arch. Ration. Mech. Anal., to appear.
- [55] M. Heinzele, A.D. Rendall and C. Ugla, *Theory of Newtonian self-gravitating stationary spherically symmetric systems*, Preprint (2004); available at [arXiv: math-ph/0408045](https://arxiv.org/abs/math-ph/0408045).
- [56] D.D. Holm, J.E. Marsden, T. Ratiu and A. Weinstein, *Nonlinear stability of fluid and plasma equilibria*, Phys. Reports **123** (1/2) (1985), 1–116.
- [57] E. Horst, *On the classical solutions of the initial value problem for the unmodified non-linear Vlasov equation I*, Math. Methods Appl. Sci. **3** (1981), 229–248.
- [58] E. Horst, *On the classical solutions of the initial value problem for the unmodified non-linear Vlasov equation II*, Math. Methods Appl. Sci. **4** (1982) 19–32.
- [59] E. Horst, *Symmetric plasmas and their decay*, Comm. Math. Phys. **126** (1990) 613–633.
- [60] E. Horst, *On the asymptotic growth of the solutions of the Vlasov–Poisson system*, Math. Methods Appl. Sci. **16** (1993), 75–85.
- [61] E. Horst and R. Hunze, *Weak solutions of the initial value problem for the unmodified non-linear Vlasov equation*, Math. Methods Appl. Sci. **6** (1984), 262–279.
- [62] H.J. Hwang, *Regularity for the Vlasov–Poisson system in a convex domain*, SIAM J. Math. Anal. **36** (2004), 121–171.
- [63] R. Illner and G. Rein, *Time decay of the solutions of the Vlasov–Poisson system in the plasma physical case*, Math. Methods Appl. Sci. **19** (1996), 1409–1413.
- [64] J. Jang, *Nonlinear instability in gravitational Euler–Poisson system for $\gamma = 6/5$* , Preprint (2005).
- [65] J. Jeans, *On the theory of star-streaming and the structure of the universe*, Mon. Not. R. Astron. Soc. **76** (1915), 70–84; **76** (1916), 552–567.

- [66] S. Klainerman and G. Staffilani, *A new approach to study the Vlasov–Maxwell system*, Commun. Pure Appl. Anal. **1** (2002), 103–125.
- [67] K.-O. Kruse and G. Rein, *A stability result for the relativistic Vlasov–Maxwell system*, Arch. Ration. Mech. Anal. **121** (1992), 187–203.
- [68] M. Kunze and A.D. Rendall, *The Vlasov–Poisson system with radiation damping*, Ann. Inst. H. Poincaré Anal. Non Linéaire **2** (2001), 857–886.
- [69] R. Kurth, *Das Anfangswertproblem der Stelldynamik*, Z. Astrophys. **30** (1952), 213–229.
- [70] R. Kurth, *A global particular solution to the initial-value problem of stellar dynamics*, Quart. Appl. Math., **36** (1978), 325–329.
- [71] L.D. Landau and E.M. Lifschitz, *Lehrbuch der Theoretischen Physik, Bd. X. Physikalische Kinetik*, Akademie-Verlag, Berlin (1983).
- [72] M. Lemou, F. Méhats and P. Raphael, *On the orbital stability of the ground states and the singularity formation for the gravitational Vlasov–Poisson system*, Preprint (2005).
- [73] R.L. Liboff, *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions*, Springer-Verlag, New York (2003).
- [74] E.H. Lieb and M. Loss, *Analysis*, Amer. Math. Soc., Providence, RI (1996).
- [75] Z. Lin, *Instability of periodic BGK waves*, Math. Res. Lett. **8** (2001), 521–534.
- [76] Z. Lin, *Nonlinear instability of periodic BGK waves for Vlasov–Poisson system*, Comm. Pure Appl. Math. **LVIII** (2005), 505–528.
- [77] Z. Lin and W.A. Strauss, *Linear stability and instability of relativistic Vlasov–Maxwell systems*, Comm. Pure Appl. Math., to appear.
- [78] Z. Lin and W.A. Strauss, *Nonlinear stability and instability of relativistic Vlasov–Maxwell systems*, Comm. Pure Appl. Math., to appear.
- [79] P.-L. Lions, *The concentration-compactness principle in the calculus of variations. The locally compact case. Part 1*, Ann. Inst. H. Poincaré Anal. Non Linéaire **1** (1984), 109–145.
- [80] P.-L. Lions and B. Perthame, *Propagation of moments and regularity for the 3-dimensional Vlasov–Poisson system*, Invent. Math. **105** (1991), 415–430.
- [81] G. Loeper, *Uniqueness of the solution to the Vlasov–Poisson system with bounded density*, Preprint (2005); available at [arXiv:math.AP/0504140](https://arxiv.org/abs/math.AP/0504140).
- [82] A. Majda, G. Majda and Y. Zheng, *Concentrations in the one-dimensional Vlasov–Poisson equations, I: Temporal development and non-unique weak solutions in the single component case*, Phys. D **74** (1994), 268–300.
- [83] A. Majda and Y. Zheng, *Existence of global weak solutions to one-component Vlasov–Poisson and Fokker–Planck–Poisson systems in one space dimension with measures as initial data*, Comm. Pure Appl. Math. **XLVII** (1994), 1365–1401.
- [84] J. Mawhin and M. Willem, *Critical Point Theory and Hamiltonian Systems*, Springer-Verlag, New York (1989).
- [85] H. Neunzert, *An Introduction to the Nonlinear Boltzmann–Vlasov Equation*, Kinetic Theories and the Boltzmann Equation, Lecture Notes in Math., Vol. 1048, Springer-Verlag, Berlin (1984).
- [86] C. Pallard, *On the boundedness of the momentum support of solutions to the relativistic Vlasov–Maxwell system*, Indiana Univ. Math. J., **54** (2005), 1395–1410.
- [87] B. Perthame, *Time decay, propagation of low moments and dispersive effects for kinetic equations*, Comm. Partial Differential Equations **21** (1996) 659–686.
- [88] B. Perthame, *Mathematical tools for kinetic equations*, Bull. Amer. Math. Soc. **41** (2004), 205–244.
- [89] K. Pfaffelmoser, *Global classical solutions of the Vlasov–Poisson system in three dimensions for general initial data*, J. Differential Equations **95** (1992), 281–303.
- [90] G. Rein, *Generic global solutions of the relativistic Vlasov–Maxwell system of plasma physics*, Comm. Math. Phys. **135** (1990), 41–78.
- [91] G. Rein, *Existence of stationary, collisionless plasmas in bounded domains*, Math. Methods Appl. Sci. **15** (1992), 365–374.
- [92] G. Rein, *Nonlinear stability for the Vlasov–Poisson system – the energy-Casimir method*, Math. Methods Appl. Sci. **17** (1994), 1129–1140.
- [93] G. Rein, *Growth estimates for the solutions of the Vlasov–Poisson system in the plasma physics case*, Math. Nachr. **191** (1998), 269–278.

- [94] G. Rein, *Flat steady states in stellar dynamics – existence and stability*, Comm. Math. Phys. **205** (1999), 229–247.
- [95] G. Rein, *Stationary and static stellar dynamic models with axial symmetry*, Nonlinear Anal. **41** (2000), 313–344.
- [96] G. Rein, *Stability of spherically symmetric steady states in galactic dynamics against general perturbations*, Arch. Ration. Mech. Anal. **161** (2002), 27–42.
- [97] G. Rein, *Reduction and a concentration-compactness principle for energy-Casimir functionals*, SIAM J. Math. Anal. **33** (2002), 896–912.
- [98] G. Rein, *Nonlinear stability of gaseous stars*, Arch. Ration. Mech. Anal. **168** (2003), 115–130.
- [99] G. Rein, *Global weak solutions to the relativistic Vlasov–Maxwell system revisited*, Commun. Math. Sci. **2** (2004), 145–158.
- [100] G. Rein and A.D. Rendall, *Global existence of solutions of the spherically symmetric Vlasov–Einstein system with small initial data*, Comm. Math. Phys. **150** (1992), 561–583.
- [101] G. Rein and A.D. Rendall, *Global existence of classical solutions to the Vlasov–Poisson system in a three-dimensional, cosmological setting*, Arch. Ration. Mech. Anal. **126** (1994), 183–201.
- [102] G. Rein and A.D. Rendall, *Compact support of spherically symmetric equilibria in non-relativistic and relativistic galactic dynamics*, Math. Proc. Cambridge Philos. Soc. **128** (2000), 363–380.
- [103] R. Robert, *Unicité de la solution faible à support compact de l'équation de Vlasov–Poisson*, C. R. Acad. Sci. Paris **324** (1997), 873–877.
- [104] O. Sánchez and J. Soler, *Orbital stability for polytropic galaxies*, Ann. Inst. H. Poincaré Anal. Non Linéaire, to appear.
- [105] J. Schaeffer, *The good, the bad, and the ugly*, Private communication (1991).
- [106] J. Schaeffer, *Global existence of smooth solutions to the Vlasov–Poisson system in three dimensions*, Comm. Partial Differential Equations **16** (1991), 1313–1335.
- [107] J. Schaeffer, *Steady states in galactic dynamics*, Arch. Ration. Mech. Anal. **172** (2004), 1–19.
- [108] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton Univ. Press, Princeton, NJ (1970).
- [109] A.A. Vlasov, Zh. Eksper. Teor. Fiz. **8** (1938), 291.
- [110] A.A. Vlasov, *Vlasov equation and plasma dispersion relation*, J. Phys. (U.S.S.R.) **9** (1945), 25.
- [111] J. Weckler, *On the initial-boundary-value problem for the Vlasov–Poisson system: Existence of weak solutions and stability*, Arch. Ration. Mech. Anal. **130** (1995), 145–161.
- [112] G. Wolansky, *On nonlinear stability of polytropic galaxies*, Ann. Inst. H. Poincaré Anal. Non Linéaire **16** (1999), 15–48.
- [113] G. Wolansky, *Static solutions of the Vlasov–Einstein system*, Arch. Ration. Mech. Anal. **156** (2001), 205–230.
- [114] P. Zhidkov, *On global solutions for the Vlasov–Poisson system*, Electron. J. Differ. Equ. **58** (2004), 1–11.

CHAPTER 6

Stochastic Representations for Nonlinear Parabolic PDEs

H. Mete Soner*

Koç University, Istanbul, Turkey

E-mail: msoner@ku.edu.tr

Contents

1. Introduction	479
Notation	480
2. Linear case: Feynman–Kac representation	480
2.1. Linear monotone semigroups	481
2.2. Zeroth-order term and forcing	484
2.3. Diffusions and parabolic PDEs	485
2.4. Initial value problems	487
2.5. Discrete Markov processes and simply coupled equations	488
2.6. Jump Markov processes and integro-differential equations	490
2.7. Dirichlet boundary conditions	492
2.8. Neumann condition and the Skorokhod problem	493
2.9. Stationary problems	494
3. Representation via controlled processes	495
3.1. Viscosity solutions	498
3.2. Optimal control of diffusion processes	502
3.3. Smooth value function and verification	507
3.4. Optimal control of jump Markov processes	509
3.5. Other type of control problems	510
3.6. Stochastic target and geometric problems	511
4. Backward representations	514
4.1. Backward stochastic differential equations	514
4.2. Second-order backward stochastic equations	515
5. Monte Carlo methods	521
5.1. The linear case	521
5.2. The semilinear case	522
5.3. The fully nonlinear case	523
Acknowledgement	524
References	524

*Member of the Turkish Academy of Sciences.

HANDBOOK OF DIFFERENTIAL EQUATIONS

Evolutionary Equations, volume 3

Edited by C.M. Dafermos and E. Feireisl

© 2007 Elsevier B.V. All rights reserved

Abstract

We discuss several different representations of nonlinear parabolic partial differential equations in terms of Markov processes. After a brief introduction of the linear case, different representations for nonlinear equations are discussed. One class of representations is in terms of stochastic control and differential games. An extension to geometric equations is also discussed. All of these representations are through the appropriate expected values of the data. Different type of representations are also available through backward stochastic differential equations. A recent extension to second-order backward stochastic differential equations allow us to represent all fully nonlinear scalar parabolic equations.

Keywords: Second-order backward stochastic differential equations, Fully nonlinear parabolic partial differential equations, Viscosity solutions, Superdiffusions, Feynman–Kac formula, BSDE, 2BSDE

MSC: 60H10, 35K55, 60H30, 60H35

1. Introduction

In this chapter we outline several connections between partial differential equations (PDE) and stochastic processes. These are extensions of Feynman–Kac-type representation of solutions to PDEs as the expected value of certain stochastic processes. Possible numerical implications of these connections are discussed as well. Although we restrict the scope of this paper to representation formulae, stochastic analysis provides much more analytical tools for PDEs. In particular, superdiffusions as developed by Dynkin [28–31] are related to nonlinear PDEs with a power-type nonlinearity, for reaction-diffusion equations interesting connections were used by Freidlin [37,38] to prove deep analytical results for these equations, also Barlow and Bass [7] study equations on fractals using random processes. Other important issues such as Martin boundaries, hypoellipticity and Malliavin calculus are not covered in these notes. Moreover, the theory of partial differential equation with stochastic forcing terms is not included. Interested readers may consult the papers by Lions and Souganidis [49,50] and by Buckhadam and Ma [17] and the references therein.

The starting point of most of our analysis is the celebrated Feynman–Kac formula [34, 42] which states that any solution of the linear heat equation

$$\frac{\partial u}{\partial t}(t, x) = \Delta u(t, x), \quad t > 0, x \in \mathbb{R}^d,$$

with initial condition $u(0, x) = f(x)$ with certain growth conditions (see Section 2) is given by

$$u(t, x) := E[f(x + \sqrt{2}W(t))],$$

where $W(\cdot)$ is the standard d -dimensional Brownian motion. This well-understood connection can be explained in several different ways. We will employ the semigroups to motivate this connection and to generalize it to more general stochastic processes. In that section we will briefly state Feynman–Kac-type formulae for several class of linear equations. The most general class of equations we will consider are second-order parabolic type integro-differential equations. Boundary value problems of Dirichlet and Neumann type are also discussed.

In Section 3 we extend these results to nonlinear equations of same type by using controlled stochastic processes. Since these equations do not always admit classical (or smooth) solutions, we will employ the theory of viscosity solutions to prove the representation formulae rigorously. The chief tool in this analysis is the dynamic programming principle which was first observed by Bellman [8]. The infinitesimal version of the dynamic programming principle is in fact gives the related partial differential equation. In this context, semigroup motivation plays an important role, as the theory of viscosity solutions is best explained through semigroups and the dynamic programming principle is in fact the semigroup property. We refer to the books by Bensoussan and Lions [9,10], Krylov [45] and Fleming and Soner [35] for more references and the historical development of the theory.

In that section we also provide a more recent representation formulae for geometric type equations. This is achieved by using a nonclassical control problem called target problems [66,67].

In this chapter we restrict our attention to only optimal control. However, these methods extend naturally to stochastic differential games. For this extension, we refer to the paper by Fleming and Souganidis [36] and Chapter 11 in the second edition of [35].

Another type of connection between PDEs and stochastic processes is given by backward stochastic differential equations (BSDEs in short). These formulae is analogous to method of characteristics for first-order equations. Indeed, initially BSDEs were studied by Bismut [11,12] then by Peng [58] as an extension of Pontryagin maximum principle which itself is an extension of characteristics. We provide a brief introduction to BSDEs and then outline a recent result of Cheridito, Soner, Touzi and Victoir [21]. This result extends the representation to formulae to all fully nonlinear, parabolic, second-order partial differential equations.

Last section is devoted to possible numerical implications of these formulae.

Notation

Let $d \geq 1$ be a natural number. We denote by $\mathcal{M}_{d,k}$ the set of all $d \times k$ matrices with real components, $\mathcal{M}^d = \mathcal{M}_{d,d}$. B' is the transpose of a matrix $B \in \mathcal{M}^d$ and $\text{Tr}[B]$ its trace. By $\mathcal{M}_{\text{inv}}^d$ we denote the set of all invertible matrices in $\mathcal{M}^d$, by $\mathcal{S}^d$ all symmetric matrices in $\mathcal{M}^d$, and by $\mathcal{S}_+^d$ all positive semidefinite matrices in $\mathcal{M}^d$. For $B, C \in \mathcal{M}^d$, we write $B \geq C$ if $B - C \in \mathcal{S}_+^d$. For $x \in \mathbb{R}^d$, we set

$$|x| := \sqrt{x_1^2 + \cdots + x_d^2}$$

and for $B \in \mathcal{M}^d$,

$$|B| := \sup_{x \in \mathbb{R}^d, |x| \leq 1} Bx.$$

Equalities and inequalities between random variables are always understood in the almost sure sense. $W(\cdot)$ is a multidimensional Brownian motion on a complete probability space $(\Omega, \mathcal{F}, P)$. For $t \geq 0$, we denote by $(\mathcal{F}_t)_{t \geq 0}$ a filtration satisfying the usual conditions and containing the filtration generated by $\{W(s)\}_{s \in [0, T]}$.

2. Linear case: Feynman-Kac representation

Let $W(\cdot)$ be the standard d -dimensional Brownian motion, a be positive constant, and f be a scalar-valued, continuous function f on $\mathbb{R}^d$ satisfying the growth condition

$$|f(x)| \leq C[1 + |x|^\alpha] \quad \forall x \in \mathbb{R}^d,$$

for some constants $C, \alpha \geq 0$. Then, the Feynman–Kac formula [34,42] states that

$$u(t, x) := E[f(x + a\sqrt{2}W(t))], \quad (2.1)$$

is the unique solution of the heat equation

$$\frac{\partial u}{\partial t}(t, x) := u_t(t, x) = a\Delta u(t, x), \quad t > 0, x \in \mathbb{R}^d,$$

together with the initial condition

$$u(0, x) = f(x), \quad x \in \mathbb{R}^d.$$

Indeed, once we know either if the function defined by the expected value is $C^{1,2}$, or if the heat equation has a smooth solution, then the above representation is an direct application of the Itô formula, Theorem 3.3 in [43]. The above growth condition is sufficient for either one of these conditions; see for instance Section 4.4, Remark 4.4 in [43]. Moreover, in this special case, polynomial growth can be weakened.

The above formula generalizes to a large class of Markov processes, linear equations and boundary problems. In this section we briefly and formally describe all these generalizations.

2.1. Linear monotone semigroups

The connection between the Markov processes and certain linear equations is now well understood and can be explained in many ways. In this chapter we will utilize semigroups to motivate this connection. The semigroup approach has the advantage that it generalizes to the nonlinear setting and it is well adapted to the theory of viscosity solutions. However, we use this approach only to motivate the connection and therefore our discussion of semigroups is only formal. In particular, we will not be precise about the domains of the operators.

In the initial discussion, we assume that the equations are defined on a metric space $\mathbb{D}$ which is equal to either $\mathbb{R}^d$ or to $\mathbb{R}^d \times \{1, 2, \dots, N\}$. Problems on bounded subsets of $\mathbb{R}^d$ are, of course, common and similar representation results are available for these equations as well. But, in this subsection, we restrict our analysis to problems defined on all of $\mathbb{R}^d$ or $\mathbb{R}^d \times \{1, 2, \dots, N\}$. Boundary problems will be discussed in the Sections 2.7 and 2.8.

Also, to simplify the presentation, we will consider PDEs that are backward in time. For these equations a terminal data at a given time T , instead of an initial data, is given. Then a solution is constructed for all times prior to T . Of course, there is a direct connection between terminal value problems and initial value problems through a simple time reversal. We perform this change for diffusion processes in the Section 2.4.

For all $t \geq 0$, let $\widehat{\mathcal{L}}_t$ be a linear operator on a subset $\widehat{\mathcal{D}}$ of $C_b(\mathbb{D})$ – bounded, scalar-valued, continuous functions on $\mathbb{R}^d$. Let $\varphi \in \widehat{\mathcal{D}}$ be a given function. For $T \geq 0$, consider the linear equation

$$-u_t(t, z) = (\widehat{\mathcal{L}}_t u(t, \cdot))(z), \quad \forall (t, z) \in (-\infty, T) \times \mathbb{D}, \quad (2.2)$$

together with final data

$$u(T, z) = \varphi(z) \quad \forall z \in \mathbb{D}. \quad (2.3)$$

Assume that for every T and $\varphi \in \widehat{\mathcal{D}}$ this equation has a unique smooth solution and let u be this unique solution. Clearly this solution depends on T and φ , but this dependence is always suppressed. Now, define a two-parameter family of operators

$$(\widehat{\mathcal{T}}_{t,T}\varphi)(z) := u(t, z), \quad \forall t \leq T, z \in \mathbb{D}.$$

By uniqueness, this family is a linear semigroup, i.e.,

$$\widehat{\mathcal{T}}_{t,T}\varphi = \widehat{\mathcal{T}}_{t,r}(\widehat{\mathcal{T}}_{r,T}\varphi) \quad \forall t \leq r \leq T. \quad (2.4)$$

Moreover, it is clear that the infinitesimal generator of this semigroup is the operator $\widehat{\mathcal{L}}_t$,

$$\lim_{h \downarrow 0} \frac{\widehat{\mathcal{T}}_{t,t+h}\varphi - \varphi}{h} = \widehat{\mathcal{L}}_t\varphi$$

for every $\varphi \in \widehat{\mathcal{D}}$.

We continue by constructing a similar semigroup using Markov processes. For this purpose, let $(\Omega, P, \mathcal{F})$ be probability space and $\{\mathcal{F}_r\}_{r \geq 0}$ be a filtration. Let $\{Z^{t,z}(r)\}_{r \geq t}$ be a $\mathbb{D}$ -valued Markov process on this probabilistic structure starting from $Z^{t,z}(t) = z$. For a continuous bounded function φ , define a two parameter semigroup by

$$\mathcal{T}_{t,T}\varphi(z) := E[\varphi(Z^{t,z}(T))] \quad \forall t \leq T.$$

Formally, the Markov property of Z implies that $\mathcal{T}_{t,T}$ satisfies (2.4). Indeed, for $t \leq r \leq T$,

$$\mathcal{T}_{t,r}(\mathcal{T}_{r,T}\varphi)(z) = E[(\mathcal{T}_{r,T}\varphi)(Z^{t,z}(r))] = E(E[\varphi(Z^{r,Z^{t,z}(r)}(T))]).$$

Since, by the Markov property of the process,

$$Z^{r,Z^{t,z}(r)}(T) = Z^{t,z}(T), \quad (2.5)$$

we have

$$\mathcal{T}_{t,r}(\mathcal{T}_{r,T}\varphi)(z) = E(E[\varphi(Z^{t,z}(T))]) = \mathcal{T}_{t,T}\varphi(z).$$

Hence, $\mathcal{T}_{t,T}$ is a two-parameter semigroup. Following the theory of semigroups, the infinitesimal generator

$$\mathcal{L}_t\varphi := \lim_{t' \rightarrow t, h \downarrow 0} \frac{\mathcal{T}_{t,t+h}\varphi - \varphi}{h} \quad (2.6)$$

exists for $\varphi \in \mathcal{D} \subset C_b(\mathbb{D})$, for some subset $\mathcal{D}$ (see Section 5.1 in [43]).

Now, it is clear that to have a Feynman–Kac representation for the equations (2.2) and (2.3), we need to construct a Markov process whose infinitesimal generator $\mathcal{L}_t$ agrees with the linear operator $\widehat{\mathcal{L}}_t$ that appears in (2.2). To see this connection formally, suppose $\mathcal{L}_t = \widehat{\mathcal{L}}_t$. Fix T and φ , and set

$$u(t, z) := \mathcal{T}_{t,T}\varphi(z) = E[\varphi(Z^{t,z}(T))]. \quad (2.7)$$

We formally claim that u solves (2.2). Indeed, assume that u is smooth in the sense that it has a continuous time derivative and $u(t, \cdot) \in \mathcal{D}$ for all $t < T$. Then, by the semigroup property,

$$\begin{aligned} -u_t(t, z) &= \lim_{h \downarrow 0} \frac{u(t-h, z) - u(t, z)}{h} \\ &= \lim_{h \downarrow 0} \frac{\mathcal{T}_{t-h,T}\varphi(x) - u(t, z)}{h} \\ &= \lim_{h \downarrow 0} \frac{\mathcal{T}_{t-h,t}(u(t, \cdot))(z) - u(t, z)}{h} \\ &= \mathcal{L}_t(u(t, \cdot))(z). \end{aligned}$$

Hence, u defined by (2.7) solves (2.2). The terminal value (2.3) follows from the definition of u . Hence, if equation (2.2) together with (2.3) has a unique “smooth” solution, then it must be given by (2.7). However, in most cases we can prove this representation directly and obtain uniqueness as a by product of the representation.

In the example of a Brownian motion, it is well known that the infinitesimal generator is the Laplacian and therefore we have the representation (2.1) for the heat equation. In the same spirit, we can prove representation results for a large class of linear equations including equations with nonlocal terms. However, the semigroup generated by Markov processes are monotone and this puts a certain restriction on the operators $\mathcal{L}_t$ that are the infinitesimal generators of the semigroups constructed by Markov processes. Indeed, the monotonicity of the stochastic semigroup (2.7) is a direct consequence of the definition and stated as

$$\varphi \leq \psi \implies \mathcal{T}_{t,r}\varphi \leq \mathcal{T}_{t,r}\psi. \quad (2.8)$$

Suppose that $\varphi, \psi \in \mathcal{D}$ and there exists $z_0 \in \mathbb{D}$ such that

$$0 = (\varphi - \psi)(z_0) = \max_{\mathbb{D}}(\varphi - \psi). \quad (2.9)$$

Then $\varphi(z_0) = \psi(z_0)$, $\varphi \leq \psi$ and by monotonicity, $\mathcal{T}_{t,r}\varphi \leq \mathcal{T}_{t,r}\psi$ for every $t \leq r$. By (2.6),

$$\mathcal{L}_t\varphi(z_0) = \lim_{h \downarrow 0} \frac{\mathcal{T}_{t,t+h}\varphi(z_0) - \varphi(z_0)}{h} \leq \lim_{h \downarrow 0} \frac{\mathcal{T}_{t,t+h}\psi(z_0) - \psi(z_0)}{h} = \mathcal{L}_t\psi(z_0).$$

Hence we proved that for every t , any infinitesimal generator $\mathcal{L}_t$ of a Markov process satisfies the *maximum principle*: For any φ, ψ in the domain of $\mathcal{L}_t$, and $z_0 \in \mathbb{D}$ satisfying (2.9), we have

$$\mathcal{L}_t \varphi(z_0) \leq \mathcal{L}_t \psi(z_0). \quad (2.10)$$

This is essentially the only important restriction for equation (2.2) to have a Feynman–Kac-type representation. Also, it is important to note that the maximum principle is the crucial property for the development of viscosity solutions as well. Moreover, the maximum principle can be directly extended to nonlinear operators and this extension will be discussed in the preceding section.

To see the importance of the maximum principle, let us consider the example of a partial differential operator. So suppose that $\mathbb{D} = \mathbb{R}^d$ and $\mathcal{L}_t$ be given by

$$\mathcal{L}_t \varphi(x) = H(t, x, \varphi(x), D\varphi(x), \dots, D^k \varphi(x))$$

for some given function H . Then, by calculus, we see that $\mathcal{L}_t$ has maximum principle if and only if $k = 2$ and

$$H(t, x, u, p, B + B') \leq H(t, x, u, p, B) \quad \forall B' \geq 0. \quad (2.11)$$

(Here and in the rest of the chapter, for symmetric matrices inequalities are understood in the sense of quadratic forms.) This property means that the corresponding equation is a second-order (possibly degenerate) parabolic equation. These equations are related to diffusion processes that will be discussed in Section 2.3. There are nonlocal operators that have maximum principle and some examples will be discussed in Sections 2.5 and 2.6.

Also, the infinitesimal generators of Markov processes, again by definition, are translation invariant. Indeed, for any φ in the domain of $\mathcal{L}_t$ and a constant β , $T_{t,T}(\varphi + \beta) = (T_{t,T}\varphi) + \beta$. Hence

$$\mathcal{L}_t(\varphi + \beta) = \lim_{h \downarrow 0} \frac{T_{t,t+h}(\varphi + \beta) - (\varphi + \beta)}{h} = \lim_{h \downarrow 0} \frac{T_{t,t+h}\varphi - \varphi}{h} = \mathcal{L}_t\varphi,$$

and therefore, the infinitesimal operators of Markov processes do not contain any zeroth-order terms. However, with a minor modification in the definition of the semigroup, a zeroth-order term and a forcing function can be included in the theory. This is the subject of the next subsection.

2.2. Zeroth-order term and forcing

Let $\{Z^{t,z}(s)\}_{s \geq t}$ be as in the previous subsection. To include a term $r(t, z)u(t, z) + h(t, z)$ to equation (2.2), we modify the Markov semigroup as follows. Define random variables,

$$B(t, T; z) := \exp\left(\int_t^T r(s, Z^{t,z}(s)) ds\right)$$

and

$$H(t, T; z) := \int_t^T B(t, s; z) h(s, Z^{t,z}(s)) ds.$$

For a continuous bounded function φ , define a two parameter semigroup by

$$\mathcal{T}_{t,T}\varphi(z) := E[B(t, T; z)\varphi(Z^{t,z}(T)) + H(t, T; z)] \quad \forall t \leq T.$$

To prove the semigroup property, for $t \leq r \leq T$, observe that

$$\begin{aligned} \mathcal{T}_{t,r}(\mathcal{T}_{r,T}\varphi)(z) &= E[B(t, r; z)(\mathcal{T}_{r,T}\varphi)(Z^{t,z}(r)) + H(t, r; z)] \\ &= E(B(t, r; z)E[B(r, T; Z^{t,z}(r))\varphi(Z^{r,Z^{t,z}(r)}(T)) \\ &\quad + H(r, T; Z^{t,z}(r))] + H(t, r; z)). \end{aligned}$$

By the Markov property of $Z(\cdot)$ (or equivalently (2.5)),

$$\begin{aligned} B(t, T; z) &= B(t, r; z)B(r, T; Z^{t,z}(r)), \\ H(t, T; z) &= B(t, r; z)H(r, T; Z^{t,z}(r)) + H(t, r; z). \end{aligned}$$

Hence, we have

$$\mathcal{T}_{t,r}(\mathcal{T}_{r,T}\varphi)(z) = E(E[B(t, T; z)\varphi(Z^{t,z}(T)) + H(t, r; z)]) = \mathcal{T}_{t,T}\varphi(z).$$

Now the infinitesimal generator of this semigroup is given by

$$\begin{aligned} \tilde{\mathcal{L}}_t\varphi(z) &= \lim_{h \downarrow 0} \frac{\mathcal{T}_{t,t+h}\varphi(z) - \varphi(z)}{h} \\ &= \lim_{h \downarrow 0} \frac{1}{h} E[B(t, t+h; z)\varphi(Z^{t,z}(t+h)) - \varphi(Z^{t,z}(t+h))] \\ &\quad + \lim_{h \downarrow 0} \frac{E[H(t, t+h; z)]}{h} + \lim_{h \downarrow 0} \frac{1}{h} (E[\varphi(Z^{t,z}(t+h))] - \varphi(z)) \\ &= r(t, z)\varphi(z) + h(t, z) + \mathcal{L}_t\varphi(z). \end{aligned}$$

2.3. Diffusions and parabolic PDEs

In this section, $\mathbb{D} = \mathbb{R}^d$ and $Z = X$ is diffusion process satisfying the stochastic differential equation (SDE),

$$dX(s) = \mu(s, X(s)) ds + \sigma(s, X(s)) dW(s), \quad \forall s > t, \quad (2.12)$$

with initial condition

$$X(t) = x. \quad (2.13)$$

We assume that

$$\mu : \mathbb{R}^+ \times \mathbb{R}^d \mapsto \mathbb{R}^d, \quad \sigma : \mathbb{R}^+ \times \mathbb{R}^d \mapsto \mathcal{M}^{d,k}$$

are given functions satisfying usual conditions (cf. [43]) and $W(\cdot)$ is a $\mathbb{R}^k$ -valued Brownian motion. Then the infinitesimal generator of this process is

$$\mathcal{L}_t = \mu(t, x) \cdot \nabla + \frac{1}{2} a(t, x) : D^2, \quad (2.14)$$

where for two $d \times d$ symmetric matrices A and B ,

$$A : B := \text{trace}[AB] = \sum_{i,j=1}^d A_{i,j} B_{i,j},$$

$$a_{i,j}(t, x) = \sum_{l=1}^k \sigma_{i,l}(t, x) \sigma_{j,l}(t, x),$$

and ∇, D^2 are respectively the gradient and the Hessian with respect to the spatial variable x . Hence, the related partial differential equation is

$$\begin{aligned} -u_t(t, x) &= \mathcal{L}_t u(t, x) \\ &= \mu(t, x) \cdot \nabla u(t, x) + \frac{1}{2} a(t, x) : D^2 u(t, x) \\ &= \sum_{i=1}^d \mu_i(t, x) u_{x_i}(t, x) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(t, x) u_{x_i x_j}(t, x) \quad \text{on } (-\infty, T) \times \mathbb{R}^d. \end{aligned} \quad (2.15)$$

The connection between the diffusion processes and the above equation can be proved directly by using the Itô calculus as well. Indeed, suppose that (2.15) together with the final data (2.3) has a smooth solution u . Fix $t < T$ and $x \in \mathbb{R}^d$ and let $\{X(s) = X^{t,x}(s)\}_{s \geq t}$ be the solution of the stochastic differential equation (2.12), (2.13). By Itô formula (Theorem 3.3.6 in [43]),

$$\begin{aligned} d(u(s, X(s))) &= [u_t(s, X(s)) + (\mathcal{L}_s u(s, \cdot))(X(s))] ds \\ &\quad + \nabla u(s, X(s)) \cdot \sigma(s, X(s)) dW(s). \end{aligned}$$

By (2.15), the ds term in the above equation is zero. Hence, the process $Y(s) := u(s, X(s))$ is a local martingale. Under suitable growth conditions on u or on ∇u , we can show that

$$Y(t) = E[Y(T)] \implies u(t, x) = E[u(T, X^{t,x}(T))] = E[\varphi(X^{t,x}(T))].$$

Note that the above proof of representation using the Itô calculus has the advantage that it also proves uniqueness under some growth conditions.

A linear term and a forcing function to equation (2.15) can be added by the technique developed in Section 2.2.

2.4. Initial value problems

In this subsection, we will briefly discuss how we may translate the above results to initial value problems. Consider an initial value problem

$$v_t(t, x) = (\tilde{\mathcal{L}}_t(v(t, \cdot)))(x) \quad \text{on } (0, \infty) \times \mathbb{R}^d,$$

together with

$$v(0, x) = \varphi(x),$$

where

$$\tilde{\mathcal{L}}_t = \tilde{\mu}(t, x) \cdot \nabla + \frac{1}{2} \tilde{a}(t, x) : D^2.$$

Fix $(t, x) \in (0, \infty) \times \mathbb{R}^d$ and set $\tilde{X}(s)$ be the solution of

$$d\tilde{X}(s) = \tilde{\mu}(t-s, \tilde{X}(s)) ds + \tilde{\sigma}(t-s, \tilde{X}(s)) dW(s),$$

with initial data $\tilde{X}(0) = x$. Apply the Itô rule to the process $Y(s) := v(t-s, \tilde{X}(s))$. The result is

$$dY(s) = [-v_t(t-s, \tilde{X}(s)) + (\tilde{\mathcal{L}}_s(v(t-s, \cdot)))(\tilde{X}(s))] ds + (\cdot \cdot \cdot) dW(s).$$

Again ds term is zero by the equation and the stochastic term is a local martingale. Under suitable growth assumptions,

$$Y(0) = E[Y(t)] \implies v(t, x) = E[v(0, \tilde{X}(t))] = E[\varphi(\tilde{X}(t))].$$

This result can also be *directly* derived from the results of Section 2.3 by a time reversal. Indeed, for a given T and a solution v of the above initial value problem, set

$$u(t, x) := v(T-t, x) \quad \forall t \leq T, x \in \mathbb{R}^d,$$

so that u solves (2.15) with

$$\mu(t, x) = \tilde{\mu}(T - t, x), \quad a(t, x) = \tilde{a}(T - t, x).$$

Let $X(s) = X^{T-t, x}(s)$ be the solution of (2.12) with initial data $X^{T-t, x}(T - t) = x$. Then, $X(s) = \tilde{X}(s - T + t)$. In particular, $X(T) = \tilde{X}(t)$, and in view of the representation proved in Section 2.3,

$$v(t, x) = u(T - t, x) = E[\varphi(X(T))] = E[\varphi(\tilde{X}(t))].$$

2.5. Discrete Markov processes and simply coupled equations

In this subsection, we will first consider Markov processes on a discrete set $\{1, 2, \dots, N\}$ and then couple these processes with diffusion processes of Section 2.3.

Let $v(\cdot)$ be a Markov process on a discrete set $\Sigma := \{1, 2, \dots, N\}$. For $(i, j) \in \Sigma$ and $r \geq t \geq 0$, let

$$P_{i,j}(t, r) := P(v(r) = j | v(t) = i),$$

be the transition probabilities. Assume that the rate functions

$$p_{i,j}(t) := \frac{\partial}{\partial r} P_{i,j}(t, r)$$

exist. Since

$$\sum_j P_{i,j}(t, r) = 1, \quad P_{i,j}(t, r) \geq 0 = P_{i,j}(t, t) \quad \forall i \neq j, r \geq t,$$

we conclude that

$$p_{i,j}(t) \geq 0 \quad \text{for } i \neq j \quad \text{and} \quad p_{i,i}(t) = -\sum_{j \neq i} p_{i,j}(t).$$

Moreover, the infinitesimal generator is given by

$$\mathcal{L}_t \varphi(i) = \sum_{j=1}^N p_{i,j}(t) \varphi(j) = \sum_{j \neq i} p_{i,j}(t) [\varphi(j) - \varphi(i)].$$

Therefore, the simple difference equation

$$-u_t(t, i) = \sum_{j \neq i} p_{i,j}(t) [u(t, j) - u(t, i)] \quad \forall t \leq T, i = 1, \dots, N, \quad (2.16)$$

with terminal data

$$u(T, \cdot) = \varphi(\cdot) \quad \text{on } \Sigma,$$

has the representation

$$u(t, i) = E[\varphi(v(t)) | v(t) = i].$$

Also a linear term and a forcing term can be added to this equation by the exponential discounting technique developed in Section 2.2. Hence, the representation result covers all equations of the form

$$\begin{aligned} -u_t(t, i) &= \sum_{j=1}^N p_{i,j}(t) [u(t, j) - u(t, i)] \\ &\quad + r_i(t)u(t, i) + h_i(t) \quad \forall t \leq T, i = 1, \dots, N, \end{aligned} \quad (2.17)$$

for given functions $p_{i,j}(t) \geq 0$ for $i \neq j$ and general functions $r_i(t), h_i(t)$ without sign restrictions. Indeed, any solution of (2.17) with terminal data (2.16), has the stochastic representation,

$$u(t, i) = E[H(t, T; i) + B(t, T; i)\varphi(v(T)) | v(t) = i],$$

where as in Section 2.2,

$$\begin{aligned} B(t, T; i) &:= \exp\left(\int_t^T r_{v(s)}(s) ds\right), \\ H(t, T; i) &:= \int_t^T B(t, s; i) h_{v(s)}(s) ds. \end{aligned}$$

We will now combine the above representation with the results of Section 2.3 to obtain a representation for a simply coupled system of parabolic equations as well. Indeed, let $Z := (X, v) \in \mathbb{R}^d \times \{1, 2, \dots, N\}$ be a Markov process constructed as follows. Let $v \in \{1, 2, \dots, N\}$ be a discrete Markov process as above, and for each $i \in \{1, 2, \dots, N\}$, $X_i^{t,x}$ be a independent diffusion processes solving the SDE

$$dX_i^{t,x}(s) = \mu_i(s, X_i^{t,x}(s)) ds + \sigma_i(s, X_i^{t,x}(s)) dW(s),$$

with initial condition (2.13). Given an initial condition $t, z := (x, i)$, we start the process

$$Z^{t,z}(t) = (X(t), v(t)) = z.$$

Then, there are strictly increasing stopping times $t < \tau_1 < \tau_2 < \dots$, so that

$$\begin{aligned} v(s) &= i, \quad s \in [t, \tau_1), \\ v(s) &= i_1, \quad s \in [\tau_1, \tau_2), \quad \dots, \quad v(s) = i_N, \quad s \in [\tau_N, \tau_{N+1}). \end{aligned}$$

Given these stopping times, we define a continuous X process recursively by

$$\begin{aligned} X(s) &= X_i^{t,x}(s), & s \in [t, \tau_1], \\ X(s) &= X_{i_N}^{\tau_N, X(\tau_N)}(s), & s \in (\tau_N, \tau_{N+1}], N = 1, 2, \dots \end{aligned}$$

It is clear that $Z = Z^{t,x,i}(\cdot)$ is a Markov process with an infinitesimal generator

$$\begin{aligned} \mathcal{L}_t \varphi(x, i) &= \mu_i(t, x) \cdot \nabla \varphi(x, i) \\ &\quad + \frac{1}{2} a_i(t, x) : D^2 \varphi(x, i) + \sum_{j \neq i} p_{i,j} [\varphi(x, j) - \varphi(x, i)]. \end{aligned}$$

Therefore

$$u(t, x, i) = E[\varphi(X(T), v(T)) | Z(t) = (X(t), v(t)) = (x, i)],$$

is a solution of the coupled parabolic equation

$$-u_t(t, x, i) = (\mathcal{L}u(t, \cdot, \cdot))(x, i) \quad \forall t < T, (x, i) \in \mathbb{R}^d \times \{1, 2, \dots, N\},$$

with final data

$$u(T, x, i) = \varphi(x, i).$$

Notice that one may see the above equation as a system of coupled parabolic equations with a solution

$$v(t, \cdot) := (u(t, \cdot, 1), \dots, u(t, \cdot, N)) : \mathbb{R}^d \mapsto \mathbb{R}^N$$

for all t . However, this coupling is only through the zeroth-order terms and the coupling constants $p_{i,j}$'s are all nonnegative. For that reason, we would like to view the above equation as a scalar valued function

$$u(t, \cdot, \cdot) : \mathbb{R}^d \times \{1, 2, \dots, N\} \mapsto \mathbb{R}^1.$$

These two different point of views have been effectively used by Freidlin in his pioneering work [37] on the analysis of some reaction–diffusion equations by stochastic methods.

2.6. Jump Markov processes and integro-differential equations

In this subsection, we will consider Markov processes which solve a stochastic differential equation which is more general than the one considered in the Section 2.3. This is done by adding a stochastic integral to the standard diffusion equation (2.12). This stochastic

integral is generally independent of the Brownian motion and it is driven by a random martingale measure. In particular, this measure contains jump terms and as such they generalize all the processes considered in the previous subsections.

Precisely, let π a positive Borel measure on $\mathbb{R}^d$, called the *compensator*, satisfying

$$\int_{\mathbb{R}^d} [1 \wedge |\xi|^2] \pi(d\xi) < \infty.$$

Given this compensator measure π , there exists a random counting measure p on the Borel subsets of $\mathbb{R}^+ \times \mathbb{R}^d$ so that, for any Borel set $A \subset \mathbb{R}^+ \times \mathbb{R}^d$, $p(A)$ has a Poisson distribution with mean

$$\lambda(A) := \int_A \pi(d\xi) d\xi.$$

Moreover, $\tilde{p} := p - \lambda$ is a martingale measure. We refer to a manuscript of Skorokhod [61] for a construction of such measures, or to a recent book by Oksendall and Sulem [54].

In the manuscript of Skorokhod [61] and in the paper of Fujiwara and Kunita [39] existence and uniqueness of stochastic differential equations are also proved. Indeed, let μ, σ be as in Section 2.3 and let

$$f : \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^d$$

be a function satisfying

$$\int_{\mathbb{R}^d} [1 \wedge |f(t, x, \xi)|^2] \pi(d\xi) < \infty.$$

We assume the standard Lipschitz condition

$$\begin{aligned} & |\mu(t, x) - \mu(t, y)|^2 + |\sigma(t, x) - \sigma(t, y)|^2 \\ & + \int_{\mathbb{R}^d} |f(t, x, \xi) - f(t, y, \xi)|^2 \pi(d\xi) \leq C|x - y|^2 \end{aligned}$$

for all $t \in \mathbb{R}^1$, $x, y \in \mathbb{R}^d$ for some constant C . Then there exists a unique solution to

$$\begin{aligned} X(r) = & x + \int_t^r \mu(s, X(s)) ds + \int_t^r \sigma(s, X(s)) dW(s) \\ & + \int_t^r \int_{\mathbb{R}^d} f(s, X(s), \xi) \tilde{p}(ds \times d\xi) \quad \forall r \geq t, \end{aligned} \quad (2.18)$$

for any initial condition (t, x) , a random measure $\tilde{p}$ constructed as above and an independent standard Brownian motion $W(\cdot)$.

Moreover, the process X is a Markov process with an infinitesimal generator,

$$\mathcal{L}_t \varphi(x) = \mu(t, x) \cdot \nabla \varphi(x) + \frac{1}{2} a(t, x) : D^2 \varphi(x) + \tilde{\mathcal{L}}_t \varphi(x),$$

where the part corresponding to the random measure is given by

$$\tilde{\mathcal{L}}_t \varphi(x) = \int_{\mathbb{R}^d} [\varphi(x + f(t, x, \xi)) - \varphi(x) - f(t, x, \xi) \cdot \nabla \varphi(x)] \pi(d\xi). \quad (2.19)$$

See for instance, p. 94 of [61] or [54] for a proof. Hence, we have the stochastic representation discussed earlier for the integro-differential equation,

$$-u_t(t, x) - \mu(t, x) \cdot \nabla u(t, x) - \frac{1}{2} a(t, x) : D^2 u(t, x) - (\tilde{\mathcal{L}}_t u(t, \cdot))(x) = 0.$$

A linear term and forcing can be added as before and also with further coupling with a discrete Markov process would yield a system of integro-differential equations.

2.7. Dirichlet boundary conditions

In the previous subsections, for the ease of exposition, we restricted our discussion to problems on all of $\mathbb{R}^d$. However, with a simple absorption rule at the boundary of a given region, we can include all Dirichlet problems into this theory. In this subsection, we outline the main tools that can be used for almost all processes. Indeed, let O be an open set with smooth boundary and $\mathcal{L}_t$ be as in (2.14) and consider the boundary value problem

$$-u_t - (\mathcal{L}_t u(t, \cdot))(x) = 0 \quad \forall t < T, x \in O,$$

together with the terminal data

$$u(T, x) = \varphi(x) \quad \forall x \in O, \quad (2.20)$$

and the boundary condition,

$$u(t, x) = g(t, x) \quad \forall t < T, x \in \partial O, \quad (2.21)$$

for some given function g . Usually, we require a compatibility condition, $g(T, x) = \varphi(x)$ for all x on the boundary of O . We can now view the solution $u(t, \cdot)$ as the value of the semigroup $\mathcal{T}_{t,T}$ applied to the terminal data φ . We include the boundary conditions in the definition of the domain of this semigroup, which can be taken as

$$C_g(t, \cdot) := \{v : C(\bar{O}) \mid v(x) = g(t, x) \forall x \in \partial O\}.$$

The stochastic semigroup is defined as follows: given φ , g , T as above, and an initial condition $x \in \mathbb{R}^d$, $t < T$, let $X^{t,x}(\cdot)$ be the solution of the SDE (2.12), (2.13). Let θ be the exit time from the domain $O \times [t, T]$,

$$\theta := \inf\{s \geq t: X^{t,x}(s) \notin O\} \wedge T.$$

Set

$$\mathcal{T}_{t,T}\varphi(x) := E[\varphi(X^{t,x}(T))\chi_{\{\theta=T\}} + g(\theta, X^{t,x}(\theta))\chi_{\{\theta<T\}}].$$

Under suitable growth and regularity conditions, we can show that the two semigroups are equal to each other. Thus, we have a stochastic representation for the boundary value problem. Integro-differential equations can be dealt with similarly. However, boundary data on all of $\mathbb{R}^d \setminus O$ is needed instead of data only on ∂O .

2.8. Neumann condition and the Skorokhod problem

Neumann-type boundary conditions are included into the theory with some care. Consider the same parabolic equation with boundary condition

$$-u_\nu(t, x) := -v(t, x) \cdot \nabla u(t, x) = g(t, x) \quad \forall t < T, x \in \partial O, \quad (2.22)$$

for some given function g and a given unit vector field $v(t, x)$. We require that

$$v(t, x) \cdot n(x) > 0, \quad (2.23)$$

where $n(x)$ is unit *inward normal* to the boundary ∂O at $x \in \partial O$. To obtain a representation we use the local time on ∂O . Indeed, we modify the SDE (2.12) in the following way. Given an initial condition $x \in O$ and $t < T$, we look for continuous processes $X^{t,x}(\cdot)$ and $l(\cdot)$ satisfying for $s \in [t, T]$,

$$\begin{aligned} X^{t,x}(s) &= x + \int_t^s \mu(r, X^{t,x}(r)) dr + \int_t^s \sigma(r, X^{t,x}(r)) dW(r) \\ &\quad + \int_t^r v(r, X^{t,x}(r)) dl(r) \in \overline{O}, \end{aligned} \quad (2.24)$$

$$\begin{aligned} l(s) &= \int_t^s \chi_{\{X^{t,x}(r) \in \partial O\}} dl(r), \\ l(0) &= 0 \quad \text{and} \quad l \text{ is nondecreasing and continuous.} \end{aligned} \quad (2.25)$$

In the literature the solution $X^{t,x}(\cdot)$ called the *reflected* diffusion process, l is the *local time* and the above set of equations are called the *Skorokhod problem*. Under the usual Lipschitz conditions on μ , σ , v , and smoothness assumption on the boundary ∂O , the Skorokhod problem has a unique solution. This and more was proved by Lions and Sznitman [51].

We also refer to Lions [47] for the connection to partial differential equations and viscosity solutions.

The only difference between (2.24) and (2.12) is the last dl integral and the important requirement that $X^{t,x}(s) \in \bar{O}$ for all s . Notice that (2.25) ensures that dl increases only when the diffusion processes $X^{t,x}$ is on the boundary ∂O . Hence, formally $X^{t,x}$ processes is “reflected” on the boundary ∂O in the direction $v(t, x)$. In view of the condition (2.23) and the fact that n is the inward normal, the reflection direction $v(t, x)$ points inward from $x \in \partial O$. These guarantee that the process $X^{t,x}$ takes values in $\bar{O}$.

We now define Markov stochastic semigroup,

$$u(t, x) = E \left[\varphi(X^{t,x}(T)) + \int_t^T g(s, X^{t,x}(s)) dl(s) \right].$$

We claim that any smooth solution $v \in C^{1,2}((0, T) \times \bar{O})$ of (2.15), (2.3) and (2.22) is equal to u . Indeed, let $X^{t,x}, l$ be a solution of the Skorokhod problem with initial data $X^{t,x}(t) = x$. Since l is a monotone function, X is a semimartingale and with the use of Itô's rule we obtain

$$\begin{aligned} \varphi(X^{t,x}(T)) &= v(T, X^{t,x}(T)) \\ &= v(t, x) + M(T) + \int_t^T [v_t + \mathcal{L}v](s, X^{t,x}(s)) ds \\ &\quad + \int_t^T \nabla v(s, X^{t,x}(s)) \cdot v(s, X^{t,x}(s)) dl(s), \end{aligned}$$

where M is a local martingale with $M(t) = 0$. By (2.15), the first integrand is zero, and by (2.22) the second integrand is equal to $-g(s, X^{t,x}(s))$. Also under some suitable growth conditions $E[M(T)] = 0$. We use these observations and then take the expected value. The result is $v = u$.

Once again, a linear term and a forcing function can be added into the theory as in the Section 2.2.

2.9. Stationary problems

Time homogeneous linear problems also have similar stochastic representations. Indeed, let $r(x) \geq \beta > 0$ be a given function. Let $Z^z(\cdot)$ be a time homogeneous Markov process with infinitesimal generator $\mathcal{L}$ and initial condition $Z^z(0) = z$. Given an open set O with smooth boundary, consider the boundary value problem

$$r(z)u(z) - \mathcal{L}u(z) = h(z) \quad \forall z \in O,$$

together with the boundary condition,

$$u(z) = g(z) \quad \forall z \in \mathbb{R}^d \setminus O, \quad (2.26)$$

for some given function g .

To obtain a stochastic representation, let θ be the exit time from the domain O ,

$$\theta^z := \inf\{s \geq 0: Z^z(s) \notin O\}.$$

Set

$$u(z) := E \left[\int_0^{\theta^z} B(s) h(Z^z(s)) ds + B(\theta^z) g(Z^z(\theta^z)) \chi_{\{\theta^z < \infty\}} \right],$$

where

$$B(t) := \exp \left(- \int_0^t r(s, Z^z(s)) ds \right).$$

Notice that due our strict positivity assumption on r , $B(s) \leq e^{-\beta s}$ for all $s \geq 0$. Therefore, the integral term in the above expression is integrable under reasonable growth assumptions on h .

For $T > 0$, define a stochastic semigroup by

$$\begin{aligned} \mathcal{T}_T \varphi(z) := E & \left[\int_0^{T \wedge \theta^z} B(s) h(Z^z(s)) ds \right. \\ & \left. + B(\theta^z) \varphi(Z^z(\theta^z)) \chi_{\{\theta^z < T\}} + B(T) g(Z^z(T)) \chi_{\{\theta^z \geq T\}} \right]. \end{aligned}$$

Time homogeneity of the Markov process implies the semigroup property,

$$\mathcal{T}_{T+S} \varphi = \mathcal{T}_T (\mathcal{T}_S \varphi) \quad \forall T, S \geq 0.$$

Also, u is a fixed point of this semigroup for every T ; that is, $u = \mathcal{T}_T u$ for all $T \geq 0$. Then, under suitable growth and regularity conditions, we can show that u is the unique solution of the linear equation. Thus, we have a stochastic representation for the stationary boundary value problem as well. For partial differential equations, the boundary condition is needed only on ∂O . Neumann boundary conditions are handled as in the previous subsection.

3. Representation via controlled processes

In this section we will consider nonlinear equations of the form

$$-u_t(t, z) + \mathcal{H}(t, z, u(t, \cdot)) = 0 \quad \forall (t, x) \in (-\infty, T) \times \mathbb{D}, \quad (3.1)$$

where $\mathbb{D}$ is as before. The most general form of the nonlinearity $\mathcal{H}$ is of the form

$$\mathcal{H}(t, z, \varphi(\cdot)) := \inf_{\beta \in B} \sup_{\alpha \in A} \{ -\mathcal{L}_t^{\alpha, \beta} \varphi(z) + r(t, z, \alpha, \beta) \varphi(z) - L(t, z, \alpha, \beta) \}, \quad (3.2)$$

where for a set of parameters α, β in some control sets A and B , $\mathcal{L}_t^{\alpha, \beta}$ is the infinitesimal generator of Markov process on $\mathbb{D}$, as in Section 2.1, and r and L are given functions. These equations are related to stochastic differential games. An interesting and an important class of equations related to stochastic optimal control is obtained by taking B to be a singleton. In this case, the nonlinearity reduces to

$$\mathcal{H}(t, z, \varphi(\cdot)) := \sup_{\alpha \in A} \{-\mathcal{L}_t^\alpha \varphi(z) + r(t, z, \alpha)\varphi(z) - L(t, z, \alpha)\}. \quad (3.3)$$

In this chapter we only discuss operators of the above form. For differential games, we refer the interested reader to Chapter 11 in the *second edition* of [35].

Finally, note that the nonlinearity in (3.2) has the maximum principle as defined in Section 2.1 (cf. (2.10)). Recall that as a consequence of maximum principle all *local operators* $\mathcal{H}$ on $\mathbb{D} = \mathbb{R}^d$ of the above form must be given by

$$\mathcal{H}(t, x, \varphi(\cdot)) = H(t, x, \varphi(x), \nabla \varphi(x), D^2 \varphi(x))$$

for some given function H satisfying (2.11).

Consider the nonlinear equation (3.1). If this equation together with the terminal data (2.3) has a unique solution, then we define a nonlinear semigroup acting on the terminal data φ by

$$\widehat{T}_{t,T}(\varphi)(z) := u(t, z).$$

By uniqueness, this is a semigroup. To obtain the related stochastic semigroup, we consider $\mathcal{H}$ as in (3.3).

Let $\mathcal{A}$ be the set of all bounded, progressively measurable processes $\alpha(t) \in A$. Again as in the linear case, given an initial condition t, z and a process $\alpha(\cdot) \in \mathcal{A}$, consider a class of processes $Z^{t,z,\alpha(\cdot)}$. We assume that, for every fixed $\alpha \in A$, the infinitesimal generator of the processes $Z^{t,z,\tilde{\alpha}}$ is equal to $\mathcal{L}_t^\alpha$, where $\tilde{\alpha}$ is process which is equal to the constant α everywhere. Define the *value function* v by

$$v(t, z) := \inf_{\alpha \in \mathcal{A}} (\mathcal{J}_{t,T}^{\alpha(\cdot)} \varphi)(z), \quad (3.4)$$

where with $Z = Z^{t,z,\alpha(\cdot)}$,

$$\begin{aligned} (\mathcal{J}_{t,T}^{\alpha(\cdot)} \varphi)(z) &:= E \left[\int_t^T B(s) L(s, Z(s), \alpha(s)) ds + B(T) \varphi(Z(T)) \right], \\ B(s) &= B(t, s; z, \alpha(\cdot)) = \exp \left(- \int_t^s r(s', Z(s'), \alpha(s')) ds \right). \end{aligned}$$

Bellman's dynamic programming in this context states that, for any stopping time

$\theta \in [t, T]$,

$$\begin{aligned} v(t, z) &= \inf_{\alpha(\cdot)} \left\{ E \left[\int_t^\theta B(s) L(s, Z(s), \alpha(s)) \, ds + B(\theta) v(\theta, Z(\theta)) \right] \right\} \\ &= \inf_{\alpha(\cdot)} \{ (\mathcal{J}_{t, \theta}^{\alpha(\cdot)} v(\theta, \cdot))(z) \}. \end{aligned}$$

We refer to [65] for a general abstract proof of dynamic programming under some structural assumptions. The crucial structure needed to prove the above result is the additive structure given by the Markov assumption on the process Z and the fact that

$$B(t, T; z, \alpha(\cdot)) = B(t, \theta; z, \alpha(\cdot)) B(\theta, T; Z^{t, z, \alpha(\cdot)}(\theta), \alpha(\cdot)).$$

Now we define the stochastic semigroup by

$$(\mathcal{T}_{t, T} \varphi) = v(t, \cdot) = \inf_{\alpha \in \mathcal{A}} (\mathcal{J}_{t, T}^{\alpha(\cdot)} \varphi),$$

so that the dynamic programming principle implies that, for any stopping time $\theta \in [t, T]$,

$$(\mathcal{T}_{t, T} \varphi) = (\mathcal{T}_{t, \theta} u(\theta, \cdot)) = (\mathcal{T}_{t, \theta} [\mathcal{T}_{\theta, T} \varphi]).$$

Hence, dynamic programming implies that $\mathcal{T}_{t, T}$ is a semigroup. Indeed, the dynamic programming principle and the semigroup properties are essentially equivalent.

It now remains to show that the infinitesimal generator of this stochastic semigroup agrees with $\mathcal{H}$ given in (3.3). We show this connection only formally here. We then introduce the theory of viscosity solutions of Crandall and Lions to prove it rigorously. Indeed, we need to compute the following limit,

$$\begin{aligned} &\lim_{\theta \downarrow t} \frac{(\mathcal{T}_{t, \theta} \varphi)(z) - \varphi(z)}{\theta - t} \\ &= \lim_{\theta \downarrow t} \inf_{\alpha(\cdot) \in \mathcal{A}} \left[\frac{E \int_t^\theta B(s) L(s, Z(s), \alpha(s)) \, ds}{\theta - t} + \frac{E(B(\theta) \varphi(Z(\theta)) - \varphi(z))}{\theta - t} \right]. \end{aligned}$$

Now, formally, assume that we may interchange the order of limit and infimum. Also, again formally, assume that the infimum can be taken only over all $\bar{\alpha}$, where $\bar{\alpha}(s) = \alpha$ for all s . Then, formally,

$$\lim_{\theta \downarrow t} \frac{(\mathcal{T}_{t, \theta} \varphi)(z) - \varphi(z)}{\theta - t} = \inf_{\alpha \in \mathcal{A}} \lim_{\theta \downarrow t} [\mathcal{J}_1(t, \theta, \alpha)(z) + (\mathcal{J}_2(t, \theta, \alpha) \varphi)(z)],$$

where

$$\mathcal{J}_1(t, \theta, \alpha)(z) = \frac{E \int_t^\theta B(s) L(s, Z(s), \alpha) \, ds}{\theta - t}$$

and

$$(\mathcal{J}_2(t, \theta, \alpha)\varphi)(z) = \frac{E[B(\theta)\varphi(Z(\theta)) - \varphi(z)]}{\theta - t}.$$

Since the infinitesimal generator of the controlled process Z with control process $\bar{\alpha}$ is assumed to be $\mathcal{L}_t^\alpha$, the limit of $\mathcal{J}_2$ is equal to $\mathcal{L}_t^\alpha\varphi(z) - r(t, z, \alpha)\varphi(z)$. Also, by the continuity of the processes, the limit of $\mathcal{J}_1$ is equal to $L(t, z, \alpha)$. Hence, formally, we compute the infinitesimal generator is

$$\begin{aligned} \lim_{\theta \downarrow t} \frac{(\mathcal{T}_{t, \theta}\varphi)(z) - \varphi(z)}{\theta - t} &= \inf_{\alpha \in A} \{ \mathcal{L}_t^\alpha\varphi(z) - r(t, z, \alpha)\varphi(z) + L(t, z, \alpha) \} \\ &= -\mathcal{H}(t, z, \varphi(\cdot)). \end{aligned}$$

Hence, we have shown that the semigroup $\widehat{\mathcal{T}}_{t, T}$ related to the PDE and the stochastic semigroup $\mathcal{T}_{t, T}$ have the same infinitesimal generator. Therefore, as in the linear case, if the PDE has unique solution in a certain class containing v then, the value function v is the unique solution of the PDE.

The main focus of the preceding subsections, is to make the above calculations rigorous and to extend these results to general nonlinearities.

3.1. Viscosity solutions

This subsection follows very closely [35].

Let $\mathbb{D}$ be closed subset of a Banach space and $\mathcal{C}$ be a collection of functions on $\mathbb{D}$ which is closed under addition, i.e.,

$$\phi, \psi \in \mathcal{C} \Rightarrow \phi + \psi \in \mathcal{C}.$$

As in the previous sections, the main object of our analysis is a two parameter family of operators $\{\mathcal{T}_{t, r}: t \leq r \leq T\}$ with the common domain $\mathcal{C}$. In the applications the exact choice of $\mathcal{C}$ is not important. However, when $\mathbb{D}$ is compact, we will require that $\mathcal{C}$ contains $C(\mathbb{D})$. For noncompact Σ , additional conditions are often imposed. Indeed, in most of our examples, we will require that $\mathcal{C}$ contains $\mathcal{M}(\mathbb{D}) \cap C_p(\mathbb{D})$ ($\mathcal{M}(\mathbb{D})$ is set of all real-valued functions which are bounded from below, $C_p(\mathbb{D})$ is set of all continuous, real-valued functions which are polynomially growing). We assume that $\mathcal{T}_{t, t}$ is the identity.

Next we want to state the semigroup property. When, $\mathcal{T}_{r, T}\varphi$ belongs to $\mathcal{C}$, the semigroup property is (2.4). However, $\mathcal{T}_{r, T}\varphi$ may not be in the domain. So, in general we assume that, for all $\phi, \psi \in \mathcal{C}$ and $t \leq r \leq s \leq T$,

$$\mathcal{T}_{t, r}\phi \leq \mathcal{T}_{t, s}\psi \quad \text{if } \phi \leq \mathcal{T}_{r, s}\psi, \quad (3.5)$$

$$\mathcal{T}_{t, r}\phi \geq \mathcal{T}_{t, s}\psi \quad \text{if } \phi \geq \mathcal{T}_{r, s}\psi. \quad (3.6)$$

By taking $r = s$ in (3.2) we conclude that the above conditions imply $\mathcal{T}_{t,r}$ is monotone, (2.8). Moreover, if $\mathcal{T}_{r,s}\psi \in \mathcal{C}$, by taking $\phi = \mathcal{T}_{r,s}\psi$, we obtain (2.4). So, in general, (3.5), (3.6) is a convenient way of stating monotonicity and the semigroup properties.

All the linear semigroups introduced in the previous section satisfy the above conditions. We will now give the example of a semigroup generated by stochastic optimal. This example will then be studied in detail in the next subsection.

EXAMPLE 3.1 (Controlled diffusion processes). We follow the construction of (3.4). Let A be a control set, O be an open subset of $\mathbb{R}^d$. Set $\mathbb{D} = \overline{O}$ and $\mathcal{C} = \mathcal{M}(\Sigma)$, set of all measurable functions bounded from below. Let μ, σ, L, g be functions satisfying the standard Lipschitz conditions (see [35]), i.e., for any function

$$\phi : (-\infty, T] \times \mathbb{D} \times A \mapsto \mathcal{M},$$

where $\mathcal{M}$ is any normed space (in our applications $\mathcal{M}$ is either $\mathbb{R}^d$ or the set of real matrices with usual norm), we say that ϕ satisfies the standard Lipschitz condition if ϕ is continuous in the (t, x) variables and

$$\|\phi(t, x, \alpha) - \phi(t, x', \alpha)\|_{\mathcal{M}} \leq C|x - x'| \quad \forall t \in (-\infty, T], x, x' \in \overline{O}, \alpha \in A, \quad (3.7)$$

with a constant independent of all variables. Let $(\Omega, P, \mathcal{F})$ be a probability space, $W(\cdot)$ be a standard $\mathbb{R}^k$ Brownian motion and $\{\mathcal{F}_t\}$ be the filtration satisfying the usual conditions as in [43].

Let $\mathcal{A}$ be all bounded, progressively measurable, A -valued random processes. We call $\mathcal{A}$ the set of *admissible controls*. In some applications further restrictions on the controls are needed. These can be modeled easily by introducing (t, x) depended subsets of $\mathcal{A}$. However, in that case certain conditions must be satisfied as discussed in [35,65].

Given a process $\alpha(\cdot) \in \mathcal{A}$ and an initial condition (2.13), we consider the controlled stochastic differential equation

$$dX(s) = \mu(s, X(s), \alpha(s)) ds + \sigma(s, X(s), \alpha(s)) dW(s). \quad (3.8)$$

For a given boundary function g , a running cost function L and a function ψ , set

$$\begin{aligned} (\mathcal{J}_{t,T}^{\alpha(\cdot)} \varphi)(x) = E \left[\int_t^{\theta \wedge T} B(s) L(s, X(s), \alpha(s)) ds \right. \\ \left. + B(\theta) g(\theta, X(\theta)) \chi_{\theta < r} + B(r) \psi(X(r)) \chi_{\theta \geq r} \right], \end{aligned} \quad (3.9)$$

where

$$B(T) = B(t, T; z, \alpha(\cdot)) = \exp \left(- \int_t^T r(s, Z(s), \alpha(s)) ds \right),$$

θ is the exit time of $(s, X(s))$ from $\bar{Q} = [t, T] \times \bar{O}$. The nonlinear semigroup is given by

$$(\mathcal{T}_{t,r}\psi)(x) := \inf_{\alpha(\cdot) \in \mathcal{A}} \mathcal{J}_{t,r}^{\alpha(\cdot)} \psi(x).$$

Assume that L, g and ψ are all bounded from below. Then, $\mathcal{T}_{t,r}\psi$ is also bounded from below and therefore, for every $\psi \in \mathcal{C}$, $\mathcal{T}_{t,r}\psi$ is well defined and belongs to $\mathcal{C}$. Clearly $\mathcal{T}_{t,r}$ is monotone (2.8). Also, dynamic programming for optimal control (cf. [35,65]) implies the semigroup property (2.4).

Notice that the infinitesimal generator of the controlled process is

$$\mathcal{L}_t^\alpha = \mu(t, x, \alpha) \cdot \nabla + \frac{1}{2} a(t, x, \alpha) : D^2, \quad (3.10)$$

where “ $\cdot$ ” is as before and

$$a_{i,j}(t, x, \alpha) = \sum_{l=1}^k \sigma_{i,l}(t, x, \alpha) \sigma_{j,l}(t, x, \alpha).$$

Hence, in view of the formal argument given in the introduction of this section, the related partial differential equation is (3.1) with $\mathcal{H}$ as in (3.3) with the above $\mathcal{L}_t^\alpha$.

In the next section, we will rigorously prove the connection between the dynamic programming equation and the above semigroup.

For $\psi \in \mathcal{C}$, $t \leq T$, $x \in \mathbb{D}$, set

$$v(t, x) = (\mathcal{T}_{t,T}\psi)(x). \quad (3.11)$$

In analogy with control problems, we call $v(t, x)$ the *value function*. Using the semigroup property, we conclude that the value function satisfies

$$v(t, x) = (\mathcal{T}_{t,r}v(r, \cdot))(x) \quad \forall x \in \mathbb{D}, t \leq r \leq T, \quad (3.12)$$

provided that $v(r, \cdot) \in \mathcal{C}$. This identity is just a restatement of the dynamic programming principle when the semigroup is related to an optimal control problem. Hence, we refer to (3.12) as the *(abstract) dynamic programming principle*.

Having formulated the dynamic programming principle abstractly, we proceed to derive the corresponding dynamic programming equation. Let $r = t + h$ in (3.12) for some $h > 0$ and small. Assume that $v(t + h, \cdot) \in \mathcal{C}$. Then

$$-\frac{1}{h} [(\mathcal{T}_{t+h}v(t+h, \cdot))(x) - v(t, x)] = 0, \quad (3.13)$$

for all $x \in \mathbb{D}$ and $t < t + h \leq T$. To continue even formally, we need to assume that the above quantity has a limit as $h \downarrow 0$, when v is “smooth”. So we assume that there exist an open set $\mathbb{D}' \subset \mathbb{D}$, a set of smooth functions $\mathcal{D} \subset C((-\infty, T) \times \mathbb{D}')$ and a one-parameter

family of nonlinear operators $\{\mathcal{G}_t\}_{t \leq T}$ of functions of $\mathbb{D}$, satisfying the following conditions with

$$Q = (-\infty, T) \times \mathbb{D}', \quad \varphi_t(t, x), (\mathcal{G}_t \varphi(t, \cdot))(x) \in C(Q) \quad \text{and} \quad \varphi(t, \cdot) \in \mathcal{C} \quad \forall t \leq T, \quad (3.14)$$

$$\varphi, \tilde{\varphi} \in \mathcal{D}, \lambda \geq 0 \implies \varphi + \tilde{\varphi} \in \mathcal{D}, \lambda \varphi \in \mathcal{D}, \quad (3.15)$$

$$\lim_{h \downarrow 0} \frac{1}{h} [(\mathcal{T}_{t,t+h} \varphi(t+h, \cdot))(x) - \varphi(t, x)] = \varphi_t(t, x) - (\mathcal{G}_t \varphi(t, \cdot))(x) \quad (3.16)$$

for all $\varphi \in \mathcal{D}$, $(t, x) \in Q$. We refer to the elements of $\mathcal{D}$ as *test functions* and $\mathcal{G}_t$ as the *infinitesimal generator* of the semigroup $\{\mathcal{T}_{t,r}\}$. Note that, if φ is any test function, then $\varphi(t, x)$ is defined for all $(t, x) \in (-\infty, T] \times \mathbb{D}$ even though (3.16) is required to hold only for $(t, x) \in Q$.

Like the choice of $\mathcal{C}$, the exact choice of $\mathcal{D}$ is not important. One should think of $\mathcal{D}$ as the set of “smooth” functions. For example, if $\mathbb{D}' = O$ is a bounded subset of $\mathbb{R}^d$ and $\mathbb{D} = \overline{O}$, then we require that $\mathcal{D}$ contains $C^\infty(\overline{Q})$. Indeed, this requirement will be typical when $\mathcal{G}_t$ is a partial differential operator.

In most applications, $\mathbb{D}'$ is simply the interior of $\mathbb{D}$. However, in the case of a controlled jump Markov process which is stopped after the exit from an open set $O \subset \mathbb{R}^d$, we have $\mathbb{D}' = O$, while $\mathbb{D}$ is the closure of the set that can be reached from O .

Now suppose that $v \in \mathcal{D}$ and let h go to zero in (3.13). The result is

$$-v_t(t, x) + (\mathcal{G}_t v(t, \cdot))(x) = 0, \quad (t, x) \in Q. \quad (3.17)$$

In analogy with optimal control, the above equation is called the *(abstract) dynamic programming equation*.

In general, the value function is not in $\mathcal{D}$ and therefore it is *not* a classical solution of (3.17). In that case the equation (3.17) has to be interpreted in a weaker sense. This will be the subject of viscosity solutions.

We are now in a position to give the definition of viscosity solutions in the abstract setting. This is a straightforward generalization of the original definition given by Crandall and Lions [25]. Also see Crandall, Evans and Lions [23]. Let $Q = (-\infty, T) \times \mathbb{D}'$, $\mathcal{D}$ and $\mathcal{C}$ as before.

In the below definition, we assume continuity to simplify the presentation. However, for the definition we only need the solution to be locally bounded, see [4,35].

DEFINITION 3.2 (Viscosity solutions). Let $w \in C((-\infty, T] \times \mathbb{D})$. Then

- (i) w is a *viscosity subsolution* of (3.17) in Q if for each $\varphi \in \mathcal{D}$,

$$-\varphi_t(\bar{t}, \bar{x}) + (\mathcal{G}_{\bar{t}} \varphi(\bar{t}, \cdot))(\bar{x}) \leq 0, \quad (3.18)$$

at every $(\bar{t}, \bar{x}) \in Q$ which is a maximizer of $w - \varphi$ on $(-\infty, T] \times \mathbb{D}$ with $w(\bar{t}, \bar{x}) = \varphi(\bar{t}, \bar{x})$.

- (ii) w is a *viscosity supersolution* of (3.17) in Q if for each $\varphi \in \mathcal{D}$,

$$-\varphi_t(\bar{t}, \bar{x}) + (\mathcal{G}_{\bar{t}} \varphi(\bar{t}, \cdot))(\bar{x}) \geq 0, \quad (3.19)$$

at every $(\bar{t}, \bar{x}) \in Q$ which is a minimizer of $w - \varphi$ on $(-\infty, T] \times \mathbb{D}$ with $w(\bar{t}, \bar{x}) = \varphi(\bar{t}, \bar{x})$.

(iii) w is a *viscosity solution* of (3.17) in Q if it is both a viscosity subsolution and a viscosity supersolution of (3.17) in Q .

It follows from the monotonicity and the semigroup properties and the definitions that any classical solution of (3.17) is also a viscosity solution, see for instance [35]. Another immediate consequence is the following.

THEOREM 3.3. *Assume (3.5), (3.6), (3.14)–(3.16). Suppose that the value function v defined by (3.11) is continuous. Then, v is a viscosity solution of (3.17) in Q .*

PROOF. Let $\varphi \in \mathcal{D}$ and $(\bar{t}, \bar{x}) \in Q$ be a maximizer of the difference $v - \varphi$ on $\bar{Q}$ satisfying $v(\bar{t}, \bar{x}) = \varphi(\bar{t}, \bar{x})$. Then, $\varphi \geq v$. Using (3.6) with $\phi = \varphi(r, \cdot)$ and $s = T$, we obtain for every $r \in [\bar{t}, T]$,

$$(\mathcal{T}_{\bar{t}, r} \varphi(r, \cdot))(\bar{x}) \geq (\mathcal{T}_{\bar{t}, T} \psi)(\bar{x}) = v(\bar{t}, \bar{x}) = \varphi(\bar{t}, \bar{x}).$$

Recall that by (3.14), $\varphi(r, \cdot)$ is in the domain of $\mathcal{T}_{\bar{t}, r}$. Take $r = \bar{t} + h$ and use (3.16) to arrive at

$$-\varphi_{\bar{t}}(\bar{t}, \bar{x}) + (\mathcal{G}_{\bar{t}} \varphi(\bar{t}, \cdot))(\bar{x}) = -\lim_{h \downarrow 0} \frac{1}{h} [(\mathcal{T}_{\bar{t}, \bar{t}+h} \varphi(\bar{t} + h, \cdot))(\bar{x}) - \varphi(\bar{t}, \bar{x})] \leq 0.$$

Hence (3.18) is satisfied and consequently v is a viscosity subsolution of (3.17) in Q . The supersolution property of v is proved *exactly* the same way as the subsolution property. $\square$

3.2. Optimal control of diffusion processes

In this subsection, we will prove that the value function of an optimal control problem is the unique viscosity solution of the dynamic programming equation (3.1) with $\mathcal{H}$ as in (3.3) with $\mathcal{L}^\alpha$ given by (3.10). In order to achieve this, we will define a stochastic nonlinear semigroup as in the introduction of this section and in Example 3.1. Then, we will verify the assumptions of Theorem 3.3 to show that the value function of the stochastic optimal control problem is the viscosity solution.

If the controlled Markov processes are uniformly parabolic, then there are classical solutions to the dynamic programming equation (3.17) and uniqueness is standard under natural conditions well known in the PDE literature. Combined with these PDE results, the viscosity property of the value function provides a stochastic representation. However, we do not, in general, assume the uniform parabolicity and therefore we only expect the value function to be a viscosity solution. Still in this case, there are uniqueness results for viscosity solutions (see [24, 35]) and a representation result follows. The main difference between the smooth (or equivalently the uniformly elliptic) case and the nonsmooth case is that, as for the linear problems, smooth solutions with certain growth conditions can be directly shown to be the value function. This point is further developed in the next subsection. In

this section, we consider the boundary value problems with no exponential discounting (i.e., equation with no linear term: $r \equiv 0$). However, our results easily extend to the other cases as described in Section 2.

Now let $\mathcal{T}_{t,r}$ be the semigroup defined in Example 3.1. To simplify the presentation, we take $r \equiv 0$. So that the nonlinear operators on $\mathcal{C}$ are given by

$$(\mathcal{T}_{t,r}\varphi)(x) = \inf_{\alpha(\cdot) \in \mathcal{A}} E \left\{ \int_t^{\theta \wedge r} L(s, X(s), \alpha(s)) ds + g(\theta, X(\theta)) \chi_{\theta < r} + \varphi(X(r)) \chi_{\theta \geq r} \right\},$$

where $t \leq r \leq T$, $\varphi \in \mathcal{C}$, θ is the exit time of $(s, X(s))$ from $Q = (-\infty, T] \times \bar{O}$ and $g \in C(\bar{Q})$ is a given function, which we call the *lateral boundary data*. Clearly $\mathcal{T}_{t,r}$ satisfies (3.5) and (3.6). The semigroup property however, is equivalent to the dynamic programming principle. We refer to Chapter 5 in [35] or [65] for the general structure of dynamic programming in a certain context.

To apply the results of Theorem 3.3, we also have to verify (3.16). Indeed we shall prove that (3.16) holds with $\mathbb{D} = \bar{O}$, $\mathbb{D}' = O$ (hence $Q = (-\infty, T] \times O$), $\mathcal{D} = C^{1,2}(\bar{Q})$ and the infinitesimal generator,

$$(\mathcal{G}_t\varphi)(x) = \mathcal{H}(t, x, \varphi(\cdot)), \quad (t, x) \in Q,$$

where $\mathcal{H}$ is as in (3.3) and $\mathcal{L}_t^\alpha$ as in (3.10). In view of Theorem 3.3, this result will imply that the value function is a viscosity solution of the dynamic programming equation provided that it is continuous.

Recall O is assumed to be bounded. For the unbounded case, we refer to [35].

THEOREM 3.4. *Suppose that f, σ satisfy (3.7), A is compact and g, L are continuous. Then, for every $w \in \mathcal{D}$ and $(t, x) \in Q$, we have*

$$\lim_{h \downarrow 0} \frac{1}{h} [(\mathcal{T}_{t,t+h} w(t+h, \cdot))(x) - w(t, x)] = w_t(t, x) - (\mathcal{G}_t w(t, \cdot))(x).$$

PROOF. We start with a probabilistic estimate. Let $X(\cdot)$ be the solution of (3.8) with control $\alpha(\cdot)$ and initial condition $X(t) = x \in O$. Since Q is bounded, f and σ are bounded, for any positive integer m and $h \in (0, 1]$, we have

$$\begin{aligned} E \sup_{t \leq \rho \leq t+h} |X(\rho) - x|^{2m} \\ = E \sup_{t \leq \rho \leq t+h} \left| \int_t^\rho f(s, X(s), \alpha(s)) ds + \int_t^\rho \sigma(s, X(s), \alpha(s)) dW(s) \right|^{2m} \end{aligned}$$

$$\begin{aligned}
&\leq \widehat{C}_m E \left(\int_t^{t+h} |f(s, X(s), \alpha(s))| ds \right)^{2m} \\
&\quad + \widehat{C}_m E \sup_{t \leq \rho \leq t+h} \left| \int_t^\rho \sigma(s, X(s), \alpha(s)) dW(s) \right|^{2m} \\
&\leq \widehat{C}_m \|f\|^{2m} h^{2m} + \widetilde{C}_m \|\sigma\|^{2m} h^m \leq C_m h^m,
\end{aligned} \tag{3.20}$$

where $\|\cdot\|$ denotes the sup-norm on Q and $C_m, \widehat{C}_m, \widetilde{C}_m$ are suitable constants. Set $d(x) = \text{dist}(x, \partial O)$ and recall that θ is the exit time from Q . Then for $t+h \leq T$,

$$\begin{aligned}
P(\theta \leq t+h) &\leq P\left(\sup_{t \leq \rho \leq t+h} |X(\rho) - x| \geq d(x)\right) \\
&\leq \left(E \sup_{t \leq \rho \leq t+h} |X(\rho) - x|^{2m}\right) (d(x))^{-2m} \\
&\leq \frac{C_m h^m}{(d(x))^{2m}}.
\end{aligned} \tag{3.21}$$

Fix $\alpha \in A$ and let $\bar{\alpha}(s) \equiv \alpha$. Then the definition of $\mathcal{T}_{t,t+h}$ yields

$$\begin{aligned}
I(h) &:= \frac{1}{h} [(\mathcal{T}_{t,t+h} w(t+h, \cdot))(x) - w(t, x)] \\
&\leq \frac{1}{h} E \int_t^{(t+h) \wedge \theta} L(s, X(s), \alpha) ds \\
&\quad + \frac{1}{h} E[w(t+h, X(t+h)) - w(t, x)] \chi_{\theta \geq t+h} \\
&\quad + \frac{1}{h} E[g(\theta, X(\theta)) - w(t, x)] \chi_{\theta < t+h}.
\end{aligned} \tag{3.22}$$

The estimate (3.21) with $m = 2$ yields

$$\lim_{h \downarrow 0} \frac{1}{h} P_{tx}(\theta \leq t+h) = 0$$

for every $(t, x) \in Q$. Hence

$$\lim_{h \downarrow 0} \frac{1}{h} E \int_t^{(t+h) \wedge \theta} L(s, X(s), \alpha) ds = L(t, x, \alpha)$$

and

$$\lim_{h \downarrow 0} \frac{1}{h} E[g(\theta, X(\theta)) - w(t, x)] \chi_{\theta < t+h} = 0.$$

Also, by Itô's formula

$$\begin{aligned}
 & \lim_{h \downarrow 0} \frac{1}{h} E[w(t+h, x(t+h)) - w(t, x)] \chi_{\theta \geq t+h} \\
 &= \lim_{h \downarrow 0} \frac{1}{h} E[w((t+h) \wedge \theta, X((t+h) \wedge \theta)) - w(t, x)] \\
 &= \lim_{h \downarrow 0} \frac{1}{h} E \int_t^{(t+h) \wedge \theta} [w_t(s, X(s)) + (\mathcal{L}_s^\alpha w(s, \cdot))(X(s))] ds \\
 &= (\mathcal{L}_t^\alpha w(t, \cdot))(x).
 \end{aligned}$$

Substitute the above into (3.21) to obtain

$$\limsup_{h \downarrow 0} I(h) \leq L(t, x, \alpha) + w_t(t, x) + (\mathcal{L}_t^\alpha w(t, \cdot))(x)$$

for all $\alpha \in A$. We take the infimum over α . The result is

$$\limsup_{h \downarrow 0} I(h) \leq w_t(t, x) - (\mathcal{G}_t w(t, \cdot))(x).$$

For any sequence $h_n \downarrow 0$, there exists $\alpha_n(\cdot)$ satisfying

$$\begin{aligned}
 & (\mathcal{T}_{t, t_n} w(t_n, \cdot))(x) \\
 & \geq E \left[\int_t^{\theta_n} L(s, x_n(s), \alpha_n(s)) ds + g(\theta_n, X_n(\theta_n)) \chi_{\theta_n < t_n} \right. \\
 & \quad \left. + w(t_n, X_n(t_n)) \chi_{\theta_n = t_n} \right] - (h_n)^2,
 \end{aligned}$$

where $t_n = t + h_n$, $\theta_n = \hat{\theta}_n \wedge t_n$, $X_n(\cdot)$ is the solution of (3.8), (2.13) with control α_n , and $\hat{\theta}_n$ is the exit time of $(s, X_n(s))$ from Q . Therefore

$$\begin{aligned}
 I(h_n) & \geq \frac{1}{h_n} E \int_t^{\theta_n} L(s, X_n(s), \alpha_n(s)) ds \\
 & \quad + \frac{1}{h_n} E[w(t_n, X(t_n)) - w(t, x)] \chi_{\theta_n = t_n} \\
 & \quad + \frac{1}{h_n} E[g(\theta_n, X_n(\theta_n)) - w(t, x)] \chi_{\theta_n < t_n} - h_n.
 \end{aligned} \tag{3.23}$$

The probabilistic estimate (3.21) with $m = 2$ implies that the limit of the third term is zero

and

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left| \frac{1}{h_n} E \left(\int_t^{t_n} L(t, x, \alpha_n(s)) \, ds - \int_t^{\theta_n} L(s, X_n(s), \alpha_n(s)) \, ds \right) \right| \\ & \leq \lim_{n \rightarrow \infty} \frac{1}{h_n} \left[\|L\|_\infty E(t_n - \theta_n) \right. \\ & \quad \left. + E \int_t^{t_n} |L(t, x, \alpha_n(s)) - L(s, X_n(s), \alpha_n(s))| \, ds \right]. \end{aligned} \quad (3.24)$$

Since $\bar{Q} \times A$ is compact, L is uniformly continuous. Also (3.20) implies that, for every $\delta > 0$,

$$\lim_{n \rightarrow \infty} P \left(\sup_{t \leq \rho \leq t+h_n} |X_n(\rho) - x| \geq \delta \right) = 0.$$

Therefore the uniform continuity of L and (3.20) imply that the limits in (3.24) are zero. We now use (3.21) and Itô formula to obtain

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{h_n} \left| E \left\{ [w(t_n, X(t_n)) - w(t, x)] \chi_{\theta_n=t_n} \right. \right. \\ & \quad \left. \left. - \int_t^{t_n} w_t(t, x) + (\mathcal{L}_t^{\alpha_n(s)} w(t, \cdot))(x) \, ds \right\} \right| \\ & \leq \lim_{n \rightarrow \infty} \frac{1}{h_n} \sup_\alpha |w_t(t, x) + (\mathcal{L}_t^\alpha w(t, \cdot))(x)| E(t_n - \theta_n) \\ & \quad + \lim_{n \rightarrow \infty} \frac{1}{h_n} E \int_t^{\theta_n} |(\mathcal{L}_t^{\alpha_n(s)} w(s, \cdot))(X_n(s)) - (\mathcal{L}_s^{\alpha_n(s)} w(t, \cdot))(x)| \, ds \\ & \quad + \lim_{n \rightarrow \infty} \frac{1}{h_n} E \int_t^{\theta_n} |w_t(s, X_n(s)) - w_t(t, x)| \, ds. \end{aligned}$$

Since $w \in C^{1,2}(\bar{Q})$, $\mathcal{L}_s^\alpha w(s, y)$ is a uniformly continuous function of $\bar{Q}$. As in (3.24), the dominated convergence theorem and (3.20) imply that the above limit is zero. Combine this with (3.23) and (3.24) to obtain

$$I(h_n) \geq L^n + G^n - e(n),$$

where

$$\begin{aligned} (L^n, G^n) &:= \frac{1}{h_n} \left(E \int_t^{t+h_n} L(t, x, \alpha_n(s)) \, ds, \right. \\ & \quad \left. E \int_t^{t+h_n} w_t(t, x) + (\mathcal{L}_t^{\alpha_n(s)} w(t, \cdot))(x) \, ds \right), \end{aligned}$$

and the error term $e(n)$ converges to zero as $n \rightarrow \infty$. Define a set

$$\begin{aligned}\widehat{A} = \{ (L, G) \in \mathbb{R}^2 : & L = L(t, x, \alpha), \\ & G = w_t(t, x) + (\mathcal{L}_t^\alpha w(t, \cdot))(x) \text{ for some } \alpha \in A \}.\end{aligned}$$

Then $(L^n, G^n) \in \overline{\text{co}}(\widehat{A})$, where $\overline{\text{co}}$ denotes the convex, closed hull of $\widehat{A}$. Also,

$$\begin{aligned}L^n + G^n &\geq \inf \{ L + G : (L, G) \in \overline{\text{co}}(\widehat{U}) \} \\ &= \inf \{ L + G : (L, G) \in \widehat{U} \} \\ &= w_t(t, x) - (\mathcal{G}_t w(t, \cdot))(x).\end{aligned}$$

□

As in Section 3.1, let v be the value function. Then, in view of Theorem 3.3, we have the following representation result for (3.1) with nonlinearity $\mathcal{H}$ given in (3.3) with $\mathcal{L}^\alpha$ given in (3.10) and $r \equiv 0$. However, this restriction that $r \equiv 0$ can easily be removed by the techniques developed in Section 2. For that reason we state the result including the linear term $r(t, z, \alpha)v(t, x)$. For uniqueness we need the boundary conditions. It is clear that v satisfies the terminal condition (2.20). Also, under some conditions, the value function satisfies the Dirichlet boundary condition (2.21) (see Chapter 5 in [35]). Also, in degenerate cases, (2.21) may hold only in the viscosity sense. We refer the interested reader to the book of Barles [4], or Section 7.6 in [35].

COROLLARY 3.5 (Control representation for (3.1)). *Suppose that $v \in C(\overline{Q})$. Then v , a viscosity solution of the dynamic programming equation (3.1), with nonlinearity $\mathcal{H}$ given in (3.3) with the infinitesimal generator as in (3.10), i.e.,*

$$-v_t(t, x) + H(t, x, v(t, x), D(t, x), D^2 v(t, x)) = 0 \quad \text{on } (-\infty, T) \times O, \quad (3.25)$$

where

$$\begin{aligned}H(t, x, v, p, B) \\ = \sup_{\alpha \in A} \left\{ -r(t, x, \alpha)v - \mu(t, x, \alpha) \cdot p - \frac{1}{2}a(t, x, \alpha) : B - L(t, x, \alpha) \right\}.\end{aligned}$$

In particular, if v satisfies (2.21) and if there is only one continuous viscosity solution of (3.25) together with (2.21), (2.20), then this solution is given as the value function of the stochastic optimal control problem.

3.3. Smooth value function and verification

In this subsection, we assume that there exists a $u \in C^{1,2}((-\infty, T] \times \overline{O})$ that solves the dynamic programming equation (3.25) together with boundary conditions (2.21), (2.20).

Then, we will show by a direct application of Itô calculus that this solution must be equal to the value of the stochastic semigroup defined in Example 3.1. This, in particular, proves uniqueness of $C^{1,2}$ solutions to equations (3.25), (2.20), (2.21).

THEOREM 3.6 (Verification). *Let $u \in C^{1,2}((-\infty, T] \times \bar{O})$ be a solution of (3.25), (2.20), (2.21). Then, for every $\alpha(\cdot) \in \mathcal{A}$,*

$$u(t, x) \leq (\mathcal{J}_{t,T}^{\alpha(\cdot)} \varphi)(x).$$

In addition, suppose that there exists an optimal control $\alpha^(\cdot) \in \mathcal{A}$ so that, for Lebesgue almost all $s \in [t, T]$,*

$$\begin{aligned} \alpha^*(s) \in \arg \min_{\alpha \in \mathcal{A}} & \left\{ r(s, X^*(s), \alpha^*(s)) v(s, X^*(s)) \right. \\ & - \mu(s, X^*(s), \alpha^*(s)) \cdot Dv(s, X^*(s)) \\ & - \frac{1}{2} a(s, X^*(s), \alpha^*(s)) : D^2 v(s, X^*(s)) \\ & \left. - L(s, X^*(s), \alpha^*(s)) \right\}, \end{aligned}$$

where $X^(s)$ is the solution of (3.8), with initial data (2.13) and control $\alpha^*(\cdot)$. Then*

$$u(t, x) = (\mathcal{J}_{t,T}^{\alpha^*(\cdot)} \varphi)(x).$$

PROOF. Fix (t, x) and $\alpha(\cdot)$. Let $X(\cdot)$ be the corresponding state process and $B(\cdot)$ as in (3.9). Apply the Itô rule to $Y(s) := B(s)u(s, X(s))$. The result is

$$\begin{aligned} dY(s) = B(s) & \left[-r(s, X(s), \alpha(s)) u(s, X(s)) + \mu(s, X(s), \alpha(s)) \cdot Du(s, X(s)) \right. \\ & \left. + \frac{1}{2} a(s, X(s), \alpha(s)) : D^2 u(s, X(s)) + u_t(s, X(s)) \right] ds \\ & + B(s) Du(s, X(s)) dW(s). \end{aligned}$$

We integrate the above on $[t, \theta]$, take the expected value and then use the equations (3.25), (2.21), (2.20). The result is

$$\begin{aligned} Y(t) &= u(t, x) \\ &= E[B(\theta)u(\theta, X(\theta))] \\ &\quad - E\left(\int_t^\theta B(s) \left[u_t(s, X(s)) - r(s, X(s), \alpha(s)) u(s, X(s)) \right. \right. \\ &\quad \left. \left. + \mu(s, X(s), \alpha(s)) \cdot Du(s, X(s)) \right] ds\right) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2} a(s, X(s), \alpha(s)) : D^2 u(s, X(s)) \Big] ds \Big) \\
& \geq E[B(\theta)u(\theta, X(\theta))] + E\left(\int_t^\theta B(s)L(s, X(s), \alpha(s)) ds\right) \\
& = (\mathcal{J}_{t,T}^{\alpha(\cdot)} \varphi)(x).
\end{aligned}$$

This proves the first part of the statement. To prove the second part, we repeat the above calculations with the control $\alpha^*(\cdot)$ and $X^*(\cdot)$. The inequality in the above calculation is now an equality and the optimality of $\alpha^*(\cdot)$ follows. $\square$

3.4. Optimal control of jump Markov processes

In this subsection, we briefly discuss the extension of the representation results for the integro-differential equations discussed in Section 2.6. We refer to the books [35,54] and the paper [2] for more information.

As in the diffusion case, we will introduce controlled diffusion equations driven by a Markov process and a random measure. Indeed, let A be a control set, and μ, σ, L be as in Section 3.2. Further, let π be a compensator measure on $\mathbb{R}^d$ and f be function satisfying

$$\sup_{\alpha \in A} \int_{\mathbb{R}^d} [1 \wedge |f(t, x, \xi, \alpha)|^2] \pi(d\xi) < \infty.$$

We also assume the standard Lipschitz condition

$$\int_{\mathbb{R}^d} |f(t, x, \xi, \alpha) - f(t, y, \xi, \alpha)|^2 \pi(d\xi) \leq C|x - y|^2$$

for all $t \in \mathbb{R}^1$, $x, y \in \mathbb{R}^d$ for some constant C . As before, let $\mathcal{A}$ be all bounded, progressively measurable, A -valued random processes. Then, given a control process $\alpha(\cdot) \in \mathcal{A}$ and initial point (t, x) , there exists a unique solution to

$$\begin{aligned}
X(T) = x & + \int_t^T \mu(s, X(s), \alpha(s)) ds + \int_t^T \sigma(s, X(s), \alpha(s)) dW(s) \\
& + \int_t^T \int_{\mathbb{R}^d} f(s, X(s), \xi, \alpha(s)) \tilde{p}(ds \times d\xi),
\end{aligned} \tag{3.26}$$

where $\tilde{p}$ is a martingale random measure with compensator π and $W(\cdot)$ is an independent standard Brownian motion.

For a constant control $\alpha(\cdot) \equiv \alpha$, the infinitesimal generator is given by

$$\mathcal{L}_t^\alpha \varphi(x) = \mu(t, x, \alpha) \cdot \nabla \varphi(x) + \frac{1}{2} a(t, x, \alpha) : D^2 \varphi(x) + \tilde{\mathcal{L}}_t^\alpha \varphi(x),$$

where the part corresponding to the random measure is as in (2.19),

$$\tilde{\mathcal{L}}_t^\alpha \varphi(x) = \int_{\mathbb{R}^d} [\varphi(x + f(t, x, \xi, \alpha)) - \varphi(x) - f(t, x, \xi, \alpha) \cdot \nabla \varphi(x)] \pi(d\xi).$$

Now we define the pay-off functional and the value function as in the diffusion case (see (3.9)),

$$\begin{aligned} (\mathcal{J}_{t,T}^{\alpha(\cdot)} \psi)(x) = E & \left[\int_t^T B(s) L(s, X(s), \alpha(s)) ds \right. \\ & \left. + B(\theta) g(\theta, X(\theta)) \chi_{\theta < T} + B(T) \psi(X(T)) \chi_{\theta \geq T} \right], \end{aligned}$$

where

$$B(T) = B(t, T; z, \alpha(\cdot)) = \exp\left(-\int_t^T r(s, Z(s), \alpha(s)) ds\right),$$

and θ is the exit time of $(s, X(s))$ from $\bar{Q} = [t, T] \times \bar{O}$. The nonlinear semigroup is given by

$$(\mathcal{T}_{t,r} \psi)(x) := \inf_{\alpha(\cdot) \in \mathcal{A}} \mathcal{J}_{t,r}^{\alpha(\cdot)} \psi(x).$$

Again it follows that $\mathcal{T}_{t,r}$ is monotone, semigroup by dynamic programming for optimal control.

We can then show that the value function is a viscosity solution of the corresponding dynamic programming equation. Such a result for jump Markov processes was first proved in [62], and then by Sayah [60]. We refer to the recent book of Oksendall and Sulem [54] for more information.

Given the form of the generator $\mathcal{L}_t^\alpha$ and the formal discussion given in the beginning of this section, the related dynamic programming equation is (3.1), with $\mathcal{H}$ in (3.3) and $\mathcal{L}_t^\alpha$ as above.

The boundary conditions are (2.20) and (2.21). But the important point to emphasize is that (2.21) holds for all x not only on ∂O but in all of $\mathbb{R}^d \setminus O$ as $X(\theta) \in \mathbb{R}^d \setminus O$.

3.5. Other type of control problems

Several other types of control problems have been studied in the literature. These problems are related to so-called quasivariational inequalities. Indeed, stopping time problems are related to obstacle problems [9]. Impulse or switching controls yield quasivariational inequalities [10]. Singular control problems allow the state processes to be discontinuities [35]. Dynamic programming equations for singular control problem are again quasivariational inequalities but with constraints on the first derivative of the solution. Equations with constraints on the second derivatives are much rare and obtained only in [20].

3.6. Stochastic target and geometric problems

In this section, we consider a special class of nonlinear parabolic equations. These equations are related to geometric flows of manifolds embedded in $\mathbb{R}^d$. The nonlinearities H that appear in these equations are, in addition to being parabolic (2.11), also geometric, i.e.,

$$H(t, x, \lambda p, B + \mu p \otimes p) = \lambda H(t, x, p, B) \quad \forall \lambda \geq 0, \mu \in \mathbb{R}^1. \quad (3.27)$$

It was shown in [66,67] that a large subclass of above nonlinearities have a stochastic representation similar to that discussed in Section 3.2. In this representation, however, a new class of control problems called *stochastic target problems* are used [64,65]. A stochastic target problem is a nonclassical control problem in which the controller tries to steer a controlled stochastic process into a given target set G by judicious choices of controls. The chief object of study is the set of all initial positions from which the controlled process can be steered into G with *probability one* in an allowed time interval. Clearly these *reachability sets* depend on the allowed time. Thus, they can be characterized by an evolution equation which is the analogue of the dynamic programming equation of stochastic optimal control.

Geometric equations express the velocity of the boundary as a possibly nonlinear function of the normal and the curvature vectors. In [65,66] it was shown that smooth solutions of these geometric equations, when exist, are equal to the reachability sets. However, as a Cauchy problem, these equations in general do not admit classical smooth solutions and a weak formulation is needed. Several such formulations were given starting with the pioneering work of Brakke [15]. Here we consider the viscosity formulation given independently by Chen, Giga and Goto [19] and by Evans and Spruck [33]. The main idea of this approach is to characterize the geometric solution as the zero level set of a continuous function. Then, this function solves a partial differential equation (3.25) with a geometric H satisfying (3.27).

The chief goal of this subsection is to give a stochastic characterization of the unique level set solutions of [19,33] in terms of the target problem. The stochastic semigroup is given by

$$v(t, x) := \inf_{\alpha(\cdot) \in \mathcal{A}} \operatorname{ess\,sup}_{\omega \in \Omega} \varphi(X_{t,x}^{\alpha(\cdot)}(T, \omega)), \quad (3.28)$$

where for initial data $(t, x) \in (-\infty, T) \times \mathbb{R}^d$, control process $\alpha(\cdot) \in \mathcal{A}$, the controlled process $\{X(s) := X_{t,x}^{\alpha(\cdot)}(s)\}_{s \geq t}$ is the solution of (3.8) and (2.13).

The following representation result is proved in [65].

THEOREM 3.7. *Suppose that the standard Lipschitz assumption (3.7) holds and that H is locally Lipschitz on $\{p \neq 0\}$. Then, v defined in (3.28) satisfies (2.20) pointwise and it is a discontinuous viscosity solution of (3.25) with*

$$H(t, x, p, B) := \sup_{v \in \mathcal{N}(t, x, p)} \left\{ -\mu(t, x, \alpha) \cdot p - \frac{1}{2} a(t, x, \alpha) : B \right\}, \quad (3.29)$$

where

$$\mathcal{N}(t, x, p) := \{\alpha \in A: \sigma(t, x, \alpha)p = 0\} \quad \text{for } p \neq 0 \quad \text{and} \quad \mathcal{N}(t, x, 0) := A. \quad (3.30)$$

We assume that $\mathcal{N}$ is non-empty. Observe that $H(t, x, p, B)$ defined above is geometric and also it is singular at $p = 0$ because $\mathcal{N}(t, x, 0) = A$.

The above theorem, in fact, follows from a more geometric result that connects the evolution equations more manifolds and stochastic target problems. In this context, the semigroup $\mathcal{T}_{t,T}$ acts on subsets of $\mathbb{R}^d$. Indeed, for a given Borel subset G of $\mathbb{R}^d$, the *target reachability set* is given by

$$\mathcal{T}_{t,T}G := v^G(t) := \{x \in \mathbb{R}^d: X_{t,x}^{\alpha(\cdot)}(T) \in G \text{ a.s. for some } \alpha(\cdot) \in \mathcal{A}\}.$$

Dynamic programming principle for these problems is proved in [65]: for all $t \leq r \leq T$,

$$\mathcal{T}_{t,T}G = \{x \in \mathbb{R}^d: X_{t,x}^{\alpha(\cdot)}(r) \in \mathcal{T}_{r,T}G \text{ a.s. for some } \alpha(\cdot) \in \mathcal{A}\}.$$

This is exactly the semigroup property

$$\mathcal{T}_{t,T}G = \mathcal{T}_{t,r}(\mathcal{T}_{r,T}G).$$

The infinitesimal generator of this semigroup can be stated purely in terms of geometric quantities such as the normal vector and second quadratic form of the set. Indeed, in [65] the characteristic functions of the reachability sets are shown to be viscosity solutions of the geometric dynamic programming equations in the sense defined in [63]. In particular, this result implies that the reachability set is included in the zero sublevel set of the solutions constructed in [19,33]. In view of the techniques developed by Barles, Soner and Souganidis [6], and [63], these purely geometric results are equivalent to Theorem 3.7. To state the main result in this direction we need the following definition:

$$K(t, z) := \{(\mu(t, x, \alpha), \sigma(t, x, \alpha)): \alpha \in A\}.$$

THEOREM 3.8. *Let the conditions of Theorem 3.7 hold. Suppose that φ is bounded and uniformly continuous, and (3.25) with $\mathcal{H}$ as in (3.29) has comparison. Let v is the unique bounded continuous viscosity solution of (3.25), (2.20). Assume further that the set $K(t, x)$ is closed and convex for all $(t, x) \in (-\infty, T] \times \mathbb{R}^d$. Then*

$$v^G(t) = \{x \in \mathbb{R}^d: v(t, x) \leq 0\}$$

with the target set

$$G := \{x \in \mathbb{R}^d: \varphi(x) \leq 0\}.$$

The proof of this theorem is a straightforward application of Theorem 3.7 and the results of [6]. Observe that the boundedness of φ is not a restriction, as one can replace φ by $\varphi(1 + |\varphi|)^{-1}$.

The stochastic target problems with jump-diffusion processes are discussed by Bouchard [13]. Also, target problems are related to forward-backward stochastic differential equations (FBSDEs) discussed in Section 4. A similar representation theorem for the special case of the codimension-one mean curvature flow was also obtained by Buckdahn, Cardaliaguet and Quincampoix [16].

We close this subsection by the important example of mean curvature flow.

EXAMPLE 3.9. Consider the example with $A = \mathcal{P}^{k,d}$ be the set of all projection matrices on $\mathbb{R}^d$ onto a hyperplane of dimension k , $\mu \equiv 0$ and $\sigma(t, x, \alpha) = \sqrt{2}\alpha$. Then, the state equation (3.8) reduces to

$$dX(s) = \sqrt{2}\alpha(s) dW(s),$$

where $W(\cdot)$ is the standard d -dimensional Brownian motion. Hence, at each time s , the controller decides on which k -dimensional space $X(\cdot)$ should diffuse. Then the related PDE has the form

$$\mathcal{H}^k(p, B) = \sup\{\alpha : B \mid \alpha \in \mathcal{P}^{k,d} \text{ and } \alpha p = 0\}.$$

This is exactly the same nonlinear function used by Ambrosio and Soner [1] to describe the weak flow of codimension $d - k$ mean curvature flow. In the special case of $k = d - 1$, any $\alpha \in \mathcal{P}^{d-1,d}$ is given by $\alpha = I - \nu \otimes \nu$ for some unit vector $\nu \in \mathbb{R}^d$. Also for such a matrix α and $p \neq 0$,

$$\alpha p = 0 \implies [I - \nu \otimes \nu]p = 0 \implies \nu = \pm \frac{p}{|p|} := \bar{p}.$$

Hence,

$$\mathcal{H}^{d-1}(p, B) = \sup\{[I - \nu \otimes \nu] : B \mid \nu = \pm \bar{p}\} = \text{trace}(B) - B\bar{p} \cdot \bar{p},$$

and equation (3.25) has the form

$$-v_t(t, x) - \Delta v(t, x) + \frac{D^2 v(t, x) \nabla v(t, x) \cdot \nabla v(t, x)}{|\nabla v(t, x)|^2} = 0.$$

This is the level set equation (in reversed time) for the mean curvature flow [19,33]. Then, in the special case of this example, results of this subsection can be stated as follows. The unique viscosity solution of the above level set equation of the mean curvature flow has the stochastic representation

$$v(t, x) := \inf_{\alpha(s) \in \mathcal{P}^{k,d}} \text{ess sup}_{\omega \in \Omega} \varphi \left(x + \int_t^T \sqrt{2}\alpha(s) dW(s) \right).$$

4. Backward representations

In this section we outline a different connection between PDEs and stochastic processes. Vaguely, this connection is analogous to the connection between ordinary differential equations and first-order PDEs through the method of characteristics. Indeed, it is first observed by Bismut [11] in his seminal work on the extension of Pontryagin maximum principle to stochastic optimal control. Pontryagin's maximum principle itself is the extension of the Hamilton–Jacobi theory of classical mechanics to deterministic optimal control and provides conditions for maximality through a set of ordinary differential equations. For stochastic optimal control, Bismut achieved this using stochastic processes. As well known, the method of characteristics and its mentioned generalizations have both initial and terminal boundary data to be satisfied. In the stochastic context, due to the adaptability conditions, this makes the problem harder. However, a deep theory is now developed through the recent works of Peng, Pardoux and others [32,53,55–57,59]. This theory known as *Backward stochastic differential equations* (BSDEs) will be outlined in the next subsection. BSDEs have a natural connection with PDEs and several numerical methods have been developed [27,52,68]. However, the PDEs connected to BSDEs are always quasilinear. Recently, Cheredito, Soner, Touzi and Victoir [21] extended this theory to cover all fully nonlinear, parabolic, second-order PDEs. This extension and the possible numerical implications are outlined below. For a more complete introduction to BSDEs we refer to the survey paper of El-Karoui, Peng and Quenez [32].

4.1. Backward stochastic differential equations

Let $X(\cdot) := X^{t,x}(\cdot)$ be the solution of (2.12), (2.13). Given real-valued, nonlinear function f and terminal data φ consider the equation

$$dY(s) = f(s, X(s), Y(s)) ds + Z(s) \cdot \sigma(s, X(s)) dW, \quad (4.1)$$

with terminal data

$$Y(T) = \varphi(X(T)). \quad (4.2)$$

The problem is to find processes $Y(\cdot)$ and $Z(\cdot)$ that are integrable and adapted to the filtration $\mathcal{F}_t$. Adaptedness condition is a serious technical condition as the given data for $Y(\cdot)$ is specified at the terminal time T . In the probabilistic literature the BSDE is defined more generally. Random f and more general $X(\cdot)$ process with $Y(\cdot)$ dependence are also considered. Here we restrict ourselves to the above framework to simplify the presentation.

Let us assume that the solution $Y(s)$ is given as a deterministic function of time and $X(s)$, i.e., assume that there is a deterministic function v so that

$$Y(s) = v(s, X(s)) \quad \forall s \in [t, T].$$

If we also assume that v is smooth, then by the Itô formula, we have

$$\begin{aligned} d[v(s, X(s))] &= [v_t(s, X(s)) + (\mathcal{L}_s v(s, \cdot))(X(s))] ds \\ &\quad + \nabla v(s, X(s)) \cdot \sigma(s, X(s)) dW(s), \end{aligned}$$

where $\mathcal{L}_t$ is as in (2.14). If σ has full rank, then equating the above equation to (4.1) yields

$$Z(s) = \nabla v(s, X(s)),$$

and v must solve

$$-v_t - \mathcal{L}_t v + f(t, x, v, \nabla v) = 0 \quad \text{on } (-\infty, T] \times \mathbb{R}^d.$$

Hence, smooth solutions of the above semilinear PDE has the representation in terms of the BSDE (4.1). Numerical implication of this connection is discussed in Section 5.2. Also the rigorous connection between the PDE and the BSDE is given in the references cited before.

4.2. Second-order backward stochastic equations

In the BSDE literature it has not been possible to consider PDEs with a nonlinear second-order term. Only quasilinear PDEs were shown to have connection with the BSDEs. This is achieved by introducing a Y, Z dependence in the dynamics of X .

In recent work [21], BSDEs were generalized by restricting the Z process to be a semimartingale. Precisely a *second-order backward stochastic differential equation* (2BSDE in short) has the X and Y equations, (2.12), (2.13) and (4.1), (4.2), and an additional equation

$$dZ(s) = a(s) ds + \Gamma(s) \sigma(s, X(s)) dW(s) \quad (4.3)$$

for some processes $a(\cdot)$ and $\Gamma(\cdot)$.

For simplicity, let us assume that $\sigma \equiv I_d$. Then, we rewrite the 2BSDE as

$$\begin{aligned} dX(s) &= \mu(s, X(s)) ds + dW(s), \\ dY(s) &= H(s, X(s), Y(s), Z(s), \Gamma(s)) ds + Z(s) \circ dX(s), \\ dZ(s) &= A(s) ds + \Gamma(s) dX(s), \\ Y(T) &= \varphi(X(T)), \quad X(t) = x, \end{aligned} \quad (4.4)$$

where H is a given function and the Fisk–Stratonovich integral $\circ$ is given by

$$Z(s) \circ dW(s) = Z(s) dW(s) + \frac{1}{2} \text{trace}[\Gamma(s)] ds.$$

Below, we will give the precise function spaces in which we look for the solutions. However, to establish the connection between the PDEs let us formally assume that there is a solution and is given by $Y(s) = v(s, X(s))$. Then, by the Itô formula (using the definition of the Fisk–Stratonovich integral),

$$d[v(s, X(s))] = v_t(s, X(s)) ds + \nabla v(s, X(s)) \circ dX(s).$$

Comparing this to the dY equation in (4.4), again we conclude that $Z(s) = \nabla v(s, X(s))$ and

$$-v_t(s, X(s)) + H(s, X(s), Y(s), Z(s), \Gamma(s)) = 0 \quad \forall s \in [t, T]. \quad (4.5)$$

Now we apply the Itô rule to $\nabla v(s, X(s))$. The result is

$$d[\nabla v(s, X(s))] = [\nabla v_t(s, X(s)) + \mathcal{L}_t \nabla v(s, X(s))] ds + D^2 v(s, X(s)) dW(s).$$

We compare this to the dZ equation in (4.4) to conclude that $\Gamma(s) = D^2 v(s, X(s))$. We substitute these into (4.5) to obtain

$$\begin{aligned} & -v_t(s, X(s)) \\ & + H(s, X(s), v(s, X(s)), \nabla v(s, X(s)), D^2 v(s, X(s))) = 0 \quad \forall s \in [t, T]. \end{aligned}$$

If the X process has full support, then we conclude that

$$-v_t(t, x) + H(t, x, \nabla v(t, x), D^2 v(t, x)) = 0 \quad \forall (t, x) \in (-\infty, T] \times \mathbb{R}^d. \quad (4.6)$$

Notice that we did not make any assumptions on H . In particular, no parabolicity is assumed. Although, the above formal calculations do not require parabolicity, the existence of a solution to 2BSDE fails without parabolicity. In the remainder of this subsection, we will give the precise definitions of the functions spaces and the assumptions needed. Then, we will state the main representation result of [21] without proof.

In addition to usual local Lipschitz conditions, we assume there are constants $C \geq 0$ and $p_1 \in [0, 1]$ such that

$$|\mu(t, x)| \leq C(1 + |x|^{p_1}), \quad (t, x) \in (-\infty, T] \times \mathbb{R}^d.$$

DEFINITION 4.1. Let $(t, x) \in (-\infty, T] \times \mathbb{R}^d$ and $(Y(\cdot), Z(\cdot), \Gamma(\cdot), A(\cdot))$ be a quadruple of $\mathcal{F}$ -progressively measurable processes on $[t, T]$ with values in $\mathbb{R}$, $\mathbb{R}^d$, $\mathcal{S}^d$ and $\mathbb{R}^d$, respectively. Then we call (Y, Z, Γ, A) a solution to the second-order backward stochastic differential equation (2BSDE) corresponding to $(X^{t,x}, H, \varphi)$ if they solve (4.4).

Equations (4.4) can be viewed as a whole family of 2BSDEs indexed by $(t, x) \in [0, T] \times \mathbb{R}^d$. We have formally argued that the solution of these equations are related to the fully nonlinear partial differential equation (4.6).

Since Z is a semimartingale, the use of the Fisk–Stratonovich integral in (4.4) means no loss of generality, but it simplifies the notation in the PDE (4.6). Alternatively, (4.4) could be written in terms of the Itô integral as

$$dY(s) = \tilde{H}(s, X^{t,x}(s), Y(s), Z(s), \Gamma(s)) ds + Z(s) \cdot dX^{t,x}(s),$$

where (recalling that σ in the X equation is taken to be the identity)

$$\tilde{H}(t, x, y, z, \gamma) = H(t, x, y, z, \gamma) + \frac{1}{2} \text{trace}[\gamma].$$

In terms of $\tilde{H}$, the PDE (4.6) can be rewritten as

$$-v_t(t, x) + \tilde{H}(t, x, v(t, x), Dv(t, x), D^2v(t, x)) - \frac{1}{2} \Delta v(t, x) = 0.$$

Finally, notice that the form of the PDE (4.6) does not depend on the functions the dynamics of the X process. So, we could restrict our attention to the case where $\mu \equiv 0$ and $\sigma \equiv I_d$, the $d \times d$ identity matrix. But the freedom to choose the dynamics of X from a more general class of diffusions provides additional flexibility in the design of the Monte Carlo schemes discussed in Section 5.

From a solution of the PDE to a solution of the 2BSDE. Assume $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a continuous function such that

$$v_t, Dv, D^2v, \mathcal{L}Dv \text{ exist and are continuous on } [0, T] \times \mathbb{R}^d,$$

and v solves the PDE (4.6) with terminal condition (2.20). Then it follows directly from Itô's formula that for each pair $(t, x) \in (-\infty, T) \times \mathbb{R}^d$, the processes

$$\begin{aligned} Y(s) &= v(s, X^{t,x}(s)), & s &\in [t, T], \\ Z(s) &= Dv(s, X^{t,x}(s)), & s &\in [t, T], \\ \Gamma(s) &= D^2v(s, X^{t,x}(s)), & s &\in [t, T], \\ A(s) &= \mathcal{L}Dv(s, X^{t,x}(s)), & s &\in [t, T], \end{aligned}$$

solve the 2BSDE corresponding to $(X^{t,x}, H, \varphi)$.

From a solution of the 2BSDE to a solution of the PDE. In all of this subsection, we assume that

$$H : (-\infty, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times S^d \rightarrow \mathbb{R} \quad \text{and} \quad \varphi : \mathbb{R}^d \rightarrow \mathbb{R}$$

are continuous functions that satisfy the following Lipschitz and growth assumptions:

(A1) For every $N \geq 1$ there exists a constant F_N such that

$$|H(t, x, y, z, \gamma) - H(t, x, \tilde{y}, z, \gamma)| \leq F_N |y - \tilde{y}|$$

for all $t \in (-\infty, T]$, $x, z \in \mathbb{R}^d$, $y, \tilde{y} \in \mathbb{R}^1$, $\gamma \in \mathcal{S}^d$ with $\max\{|x|, |y|, |\tilde{y}|, |z|, |\gamma|\} \leq N$.

(A2) There exist constants F and $p_2 \geq 0$ such that

$$|H(t, x, y, z, \gamma)| \leq F(1 + |x|^{p_2} + |y| + |z|^{p_2} + |\gamma|^{p_2})$$

for all $(t, x, y, z, \gamma) \in (-\infty, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{S}^d$.

(A3) There exist constants G and $p_3 \geq 0$ such that

$$|\varphi(x)| \leq G(1 + |x|^{p_3}) \quad \text{for all } x \in \mathbb{R}^d.$$

Admissible strategies. We fix constants $p_4, p_5 \geq 0$ and denote for all $(t, x) \in (-\infty, T] \times \mathbb{R}^d$ and $m \geq 0$ by $\mathcal{A}_m^{t,x}$ the class of all processes of the form

$$Z(s) = z + \int_t^s A(r) dr + \int_t^s \Gamma(r) dX^{t,x}(r), \quad s \in [t, T],$$

where $z \in \mathbb{R}^d$, $(A(\cdot), \Gamma(\cdot)) \in \mathbb{R}^d \times \mathcal{S}^d$ progressively measurable processes satisfying

$$\max\{|Z(s)|, |A(s)|, |\Gamma(s)|\} \leq m(1 + |X^{t,x}(s)|^{p_4}) \quad \forall s \in [t, T], \quad (4.7)$$

and

$$\begin{aligned} |\Gamma(r) - \Gamma(s)| &\leq m(1 + |X^{t,x}(r)|^{p_5} + |X^{t,x}(s)|^{p_5}) \\ &\quad \times (|r - s| + |X^{t,x}(r) - X^{t,x}(s)|) \quad \forall r, s \in [t, T]. \end{aligned} \quad (4.8)$$

Set $\mathcal{A}^{t,x} := \bigcup_{m \geq 0} \mathcal{A}_m^{t,x}$. It follows from the assumptions (A1) and (A2) on H and the condition (4.7) on Z that for all $y \in \mathbb{R}$ and $Z \in \mathcal{A}^{t,x}$, the forward SDE

$$dY(s) = f(s, X^{t,x}(s), Y(s), Z(s), \Gamma(s)) ds + Z(s) \circ dX^{t,x}(s), \quad s \in [t, T],$$

with $Y(t) = y$, has a unique strong solution $Y^{t,x,y,Z}(\cdot)$ (this can, for instance, be shown with the arguments in the proofs of Theorems 2.3, 2.4 and 3.1 in Chapter IV of Ikeda and Watanabe [40]).

Notice that $Z \in \mathcal{A}^{t,x}$ is a solution of the 2BSDE if $Y^{t,x,y,Z}(T) = \varphi(X^{t,x}(T))$.

We will show that solutions to 2BSDE in the class $\mathcal{A}^{t,x}$ has at most one solution.

Uniqueness of 2BSDE in $\mathcal{A}^{t,x}$. For our last assumption and the statement of Theorem 4.3, we need the following definition.

DEFINITION 4.2. Let $q \geq 0$.

(1) We call a function $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ a viscosity solution with growth q of the PDE (4.6) with terminal condition (2.20) if v is a viscosity solution of (4.6) on $(-\infty, T) \times \mathbb{R}^d$ such that $v^*(T, x) = v_*(T, x) = g(x)$ for all $x \in \mathbb{R}^d$ and there exists a constant C such that $|v(t, x)| \leq C(1 + |x|^q)$ for all $(t, x) \in (-\infty, T] \times \mathbb{R}^d$.

(2) We say that the PDE (4.6) with terminal condition (2.20) has *comparison with growth q* if the following holds:

If $w : (-\infty, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is lower semicontinuous and a viscosity supersolution of (4.6) on $(-\infty, T) \times \mathbb{R}^d$ and $u : (-\infty, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ upper semicontinuous and a viscosity subsolution of (4.6) on $(-\infty, T) \times \mathbb{R}^d$ such that

$$w(T, x) \geq g(x) \geq u(T, x) \quad \text{for all } x \in \mathbb{R}^d$$

and there exists a constant $C \geq 0$ with

$$w(t, x) \geq -C(1 + |x|^p) \quad \text{and} \quad u(t, x) \leq C(1 + |x|^p) \\ \text{for all } (t, x) \in (-\infty, T) \times \mathbb{R}^d,$$

then $w \geq u$ on $(-\infty, T] \times \mathbb{R}^d$.

With this definition our last assumption on H and φ is

(A4) The PDE (4.6) with terminal condition (2.20) has comparison with growth $p = \max\{p_2, p_3, p_2 p_4, p_4 + 2p_1\}$.

The following result is proved in [20].

THEOREM 4.3 (Uniqueness of 2BSDE). Assume (A1)–(A4) and that H is parabolic (2.11). For $x_0 \in \mathbb{R}^d$ suppose that the 2BSDE corresponding to (X^{0,x_0}, H, φ) has a solution with $Z^{0,x_0} \in \mathcal{A}^{0,x_0}$. Then

(i) The associated PDE (4.6) with terminal condition (2.20) has a unique viscosity solution v with growth $p = \max\{p_2, p_3, p_2 p_4, p_4 + 2p_1\}$, and v is continuous on $[0, T] \times \mathbb{R}^d$.

(ii) For all $(t, x) \in [0, T) \times \mathbb{R}^d$, there exists exactly one solution $(Y^{t,x}, Z^{t,x}, \Gamma^{t,x}, A^{t,x})$ to the 2BSDE corresponding to $(X^{t,x}, H, \varphi)$ such that $Z^{t,x} \in \mathcal{A}^{t,x}$ and

$$Y^{t,x}(s) = v(s, X^{t,x}(s)), \quad s \in [t, T],$$

where v is the unique continuous viscosity solution with growth p of (4.6) and (2.20).

REMARK 4.4. 1. Under the hypothesis of the above theorem, the solution of the 2BSDE satisfies $Y^{t,x}(t) = v(t, x)$. Hence, $v(t, x)$ can be approximated by backward simulation

of the process $(Y^{t,x}(s))_{s \in [t, T]}$. If v is C^2 , it follows from Itô's lemma that $Z^{t,x}(s) = Dv(s, X^{t,x}(s))$, $s \in [t, T]$. Then $Dv(t, x)$ can also be approximated by backward simulation. Moreover, for v is C^3 , $\Gamma^{t,x}(s) = D^2v(s, X^{t,x}(s))$ can be simulated in this way. A formal discussion of a potential numerical scheme for the backward simulation of the processes $Y^{t,x}$, $Z^{t,x}$ and $\Gamma^{t,x}$ is provided in Section 5.3.

2. We have already shown that a classical solution v of (4.6) and (2.20) and its derivatives provide a solution of the 2BSDE.

3. The parabolicity assumption (2.11) is natural from the PDE viewpoint. If H is uniformly elliptic: there exists a constant $C > 0$ such that

$$H(t, x, y, z, \gamma - B) \geq H(t, x, y, z, \gamma) + C \operatorname{Tr}[B] \quad \forall B \geq 0.$$

Then the PDE (4.6) is uniformly parabolic, and there exist general results on existence, uniqueness and smoothness of solutions, see for instance, [44]. When H is linear in the γ variable (in particular, for the semi- and quasilinear equations discussed in Section 5.2), uniform ellipticity essentially guarantees existence, uniqueness and smoothness of solutions to the PDE (4.6) and (2.20); see for instance, Section 5.4 in [46].

4. Condition (A4) is an implicit assumption on the functions H and φ as we find it more convenient to assume comparison directly in the form (A4) instead of placing technical assumptions on H and φ which guarantee that the PDE (4.6) with terminal condition (2.20) has comparison. However, several comparison results for nonlinear PDEs are available in the literature; see for example, Crandall, Ishii and Lions [24], Fleming and Soner [35], Cabre and Caffarelli [18]. However, most results are stated for equations in bounded domains. For equations in the whole space, the critical issue is the interplay between the growth of solutions at infinity and the growth of the nonlinearity. We list some typical situations where comparison holds:

(a) *Comparison with growth 1.* Assume (A1)–(A3) and there exists a function $h: [0, \infty) \rightarrow [0, \infty)$ with $\lim_{x \rightarrow 0} h(x) = 0$ such that

$$|H(t, x, y, \alpha(x - \tilde{x}), A) - H(t, \tilde{x}, y, \alpha(x - \tilde{x}), B)| \leq h(\alpha|x - \tilde{x}|^2 + |x - \tilde{x}|)$$

for all $(t, x, \tilde{x}, y)$, $\alpha > 0$ and A, B satisfying

$$-\alpha \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \leq \begin{bmatrix} A & 0 \\ 0 & -B \end{bmatrix} \leq \alpha \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}.$$

Then it follows from Theorem 8.2 in [24] that equations of the form (4.6), (2.20) have comparison with growth 0 if the domain is bounded. If the domain is unbounded, it follows from the modifications outlined in Section 5.D of Crandall et al. [24] that (4.6) and (2.20) have comparison with growth 1.

(b) For the dynamic programming equation (3.25) related to a stochastic optimal control problem, a comparison theorem for bounded solutions is given in [35], Section 5.9, Theorem V.9.1.

(c) Many techniques in dealing with unbounded solutions were developed by Ishii [41] for first-order equations (that is, when f is independent of γ). These techniques can be

extended to second-order equations. Some related results can be found in [4,5]. In [5], in addition to comparison results for PDEs, one can also find BSDEs based on jump Markov processes.

5. Monte Carlo methods

In this section we provide a formal discussion of the numerical implications of our representation results. We start by recalling some well-known facts in the linear case. We then review some results for the semilinear and the quasilinear cases. Then, we conclude with the fully nonlinear case related to Theorem 4.3.

5.1. The linear case

In this subsection we assume that the function H is of the form

$$H(t, x, y, z, \gamma) = -\alpha(t, x) - \beta(t, x)y - \mu(x) \cdot z - \frac{1}{2}a(x) : \gamma.$$

Then (4.6) is a linear parabolic equation and we discussed already that the Feynman–Kac representation has the form

$$v(t, x) = E \left[\int_t^T B_{t,s} \alpha(s, X^{t,x}(s)) ds + B_{t,T} g(X^{t,x}(T)) \right],$$

where

$$B_{t,s} := \exp \left(\int_t^s \beta(r, X^{t,x}(r)) dr \right).$$

This representation suggests a numerical approximation of the function v by means of the so-called Monte Carlo method.

(i) Given J independent copies $\{X^j, 1 \leq j \leq J\}$ of the process $X^{t,x}$, set

$$\hat{v}^{(J)}(t, x) := \frac{1}{J} \sum_{j=1}^J \int_t^T B_{t,s}^j \alpha(s, X^j(s)) ds + B_{t,T}^j g(X^j(T)),$$

where $B_{t,s}^j := \exp(\int_t^s \beta(r, X^j(r)) dr)$. Then, it follows from the law of large numbers and the central limit theorem that

$$\begin{aligned} \hat{v}^{(J)}(t, x) &\rightarrow v(t, x) \quad \text{a.s.} \quad \text{and} \\ \sqrt{J}(\hat{v}^{(J)}(t, x) - v(t, x)) &\rightarrow N(0, \rho) \quad \text{in distribution,} \end{aligned}$$

where ρ is the variance of the random variable $\int_t^T B_{t,s} \alpha(s, X^{t,x}(s)) ds + B_{t,T} g(X^{t,x}(T))$. Hence, $\hat{v}^{(J)}(t, x)$ is a consistent approximation of $v(t, x)$. Moreover, in contrast to finite differences or finite elements methods, the error estimate is of order $J^{-1/2}$, independent of the dimension d .

(ii) In practice, it is not possible to produce independent copies $\{X^j, 1 \leq j \leq J\}$ of the process $X^{t,x}$, except in trivial cases. In most cases, the above Monte Carlo approximation is performed by replacing the process $X^{t,x}$ by a suitable discrete-time approximation X^N with time step of order N^{-1} for which independent copies $\{X^{N,j}, 1 \leq j \leq J\}$ can be produced. The simplest discrete-time approximation is the following discrete Euler scheme: Set $X_t^N = x$ and for $1 \leq n \leq N$,

$$X_{t_n}^N = X_{t_{n-1}}^N + \mu(X_{t_{n-1}}^N)(t_n - t_{n-1}) + \sigma(X_{t_{n-1}}^N)(W_{t_n} - W_{t_{n-1}}),$$

where $t_n := t + n(T - t)/N$. We refer to [68] for a survey of the main results in this area.

5.2. The semilinear case

We next consider the case where H is given by

$$H(t, x, y, z, \gamma) = \varphi(t, x, y, z) - \mu(x) \cdot z - \frac{1}{2} a(x) : \gamma.$$

Then the PDE (4.6) is semilinear. We assume that the assumptions of Theorem 4.3 are satisfied. In view of the connection between Fisk–Stratonovich and Itô integration, the 2BSDE (4.4) reduces to an uncoupled forward–backward SDE (FBSDE) of the form

$$dY(s) = \varphi(s, X^{t,x}(s), Y(s), Z(s)) ds + Z(s) \cdot \sigma(X^{t,x}(s)) dW(s),$$

with terminal data $Y(T) = g(X^{t,x}(T))$ (compare to Peng [59], Pardoux and Peng [56]). For $N \geq 1$, we denote $t_n := t + n(T - t)/N$, $n = 0, \dots, N$, and we define the discrete-time approximation Y^N of Y by the backward scheme

$$Y_T^N := g(X_T^{t,x})$$

and, for $n = 1, \dots, N$,

$$Y_{t_{n-1}}^N := E[Y_{t_n}^N | X_{t_{n-1}}^{t,x}] - \varphi(t_{n-1}, X_{t_{n-1}}^{t,x}, Y_{t_{n-1}}^N, Z_{t_{n-1}}^N)(t_n - t_{n-1}), \quad (5.9)$$

$$Z_{t_{n-1}}^N := \frac{1}{t_n - t_{n-1}} (\sigma(X_{t_{n-1}}^{t,x}))^{-1} E[(W_{t_n} - W_{t_{n-1}}) Y_{t_n}^N | X_{t_{n-1}}^{t,x}]. \quad (5.10)$$

Then, we have

$$\limsup_{N \rightarrow \infty} \sqrt{N} |Y_t^N - v(t, x)| < \infty,$$

and in case that v is C^2 also,

$$\limsup_{N \rightarrow \infty} \sqrt{N} |Z_t^N - Dv(t, x)| < \infty,$$

see for instance, Bally and Pagès [3], Bouchard and Touzi [14]. The practical implementation of this backward scheme requires the computation of the conditional expectations appearing in (5.9) and (5.10). This suggests the use of a Monte Carlo approximation, as in the linear case. But at every time step, we need to compute conditional expectations based on J independent copies $\{X^j, 1 \leq j \leq J\}$ of the process $X^{t,x}$. Recently, several approaches to this problem have been developed. We refer to Bally and Pagès [3], Bouchard and Touzi [14], Lions and Regnier [48] and the references therein for the methodology and the analysis of such nonlinear Monte Carlo methods.

We refer to Chevance [22] and to a recent article by Delarue and Menozzi [26] for Monte Carlo simulations for the quasilinear case using forward–backward stochastic equations.

5.3. The fully nonlinear case

We now discuss the case of a general H as in the previous section. Let μ, σ be as in the dynamics (2.12)

$$\tilde{H}(t, x, y, z, \gamma) = H(t, x, y, z, \gamma) + \mu(t, x) \cdot z + \frac{1}{2} a(t, x) : \gamma.$$

Then, for all (t, x) the 2BSDE corresponding to $(X^{t,x}, H, \varphi)$ can be written as

$$\begin{aligned} dY(s) &= \tilde{H}(s, X^{t,x}(s), Y(s), Z(s), \Gamma(s)) ds \\ &\quad + Z(s) \cdot \sigma(s, X^{t,x}(s)) dW(s), \quad s \in [t, T], \\ dZ(s) &= A(s) ds + \Gamma(s) dX^{t,x}(s), \quad s \in [t, T], \\ Y(T) &= g(X^{t,x}(T)). \end{aligned} \tag{5.11}$$

We assume that the conditions of Theorem 4.3 hold true, so that the PDE (4.6) has a unique viscosity solution v with growth $p = \max\{p_2, p_3, p_2 p_4, p_4 + 2p_1\}$, and there exists a unique solution $(Y^{t,x}, Z^{t,x}, \Gamma^{t,x}, A^{t,x})$ to the 2BSDE (5.11) with $Z^{t,x} \in \mathcal{A}^{t,x}$.

Comparing with the backward scheme (5.9), (5.10) in the semilinear case, we suggest the following discrete-time approximation of the processes $Y^{t,x}$, $Z^{t,x}$ and $\Gamma^{t,x}$:

$$Y_T^N := g(X_T^{t,x}), \quad Z_T^N := Dg(X_T^{t,x}),$$

and, for $n = 1, \dots, N$,

$$\begin{aligned} Y_{t_{n-1}}^N &:= E[Y_{t_n}^N | X_{t_{n-1}}^{t,x}] - \varphi(t_{n-1}, X_{t_{n-1}}^{t,x}, Y_{t_{n-1}}^N, Z_{t_{n-1}}^N, \Gamma_{t_{n-1}}^N)(t_n - t_{n-1}) \\ Z_{t_{n-1}}^N &:= \frac{1}{t_n - t_{n-1}} (\sigma(X_{t_{n-1}}^{t,x}))'^{-1} E[(W_{t_n} - W_{t_{n-1}}) Y_{t_n}^N | X_{t_{n-1}}^{t,x}], \end{aligned}$$

and

$$\Gamma_{t_{n-1}}^N := \frac{1}{t_n - t_{n-1}} E[Z_{t_n}^N (W_{t_n} - W_{t_{n-1}})' | X_{t_{n-1}}^{t,x}] \sigma(X_{t_{n-1}}^{t,x})^{-1}.$$

We expect that

$$(Y_t^N, Z_t^N, \Gamma_t^N) \rightarrow (v(t, x), Dv(t, x), D^2v(t, x)) \quad \text{as } N \rightarrow \infty.$$

However, proof of the above assertion is not yet available.

Acknowledgement

This work was partly supported by the Turkish Academy of Sciences.

References

- [1] L. Ambrosio and H.M. Soner, *Level set approach to mean curvature flow in arbitrary codimension*, J. Differential Geom. **43** (1996), 693–737.
- [2] O. Alvarez and A. Tourin, *Viscosity solutions of nonlinear integro-differential equations*, Ann. Inst. H. Poincaré **13** (1996), 293–317.
- [3] V. Bally, G. Pagès, *Error analysis of the quantization algorithm for obstacle problems*, Stochastic Process. Appl. **106** (1) (2003), 1–40.
- [4] G. Barles, *Solutions de viscosité des équations de Hamilton–Jacobi*, Math. Appl., Vol. 17, Springer-Verlag, Paris (1994).
- [5] G. Barles, R. Buckdahn and E. Pardoux, *Backward stochastic differential equations and integral-partial differential equations*, Stoch. Stoch. Rep. **60** (1/2) (1997), 57–83.
- [6] G. Barles, H.M. Soner and P.E. Souganidis, *Front propagation and phase field theory*, SIAM J. Control Optim. **31** (2) (1993), 439–469.
- [7] M.T. Barlow and R.F. Bass, *Brownian motion and harmonic analysis on Sierpinski carpets*, J. Canad. Math. **51** (4) (1999), 673–744.
- [8] R. Bellman, *Dynamic Programming*, Princeton Univ. Press, Princeton, NJ (1957).
- [9] A. Bensoussan and J.-L. Lions, *Applications des inéquations variationnelles en contrôle stochastique*, Dunod, Paris (1978).
- [10] A. Bensoussan and J.-L. Lions, *Contrôle impulsionnel et inéquations quasi-variationnelles*, Dunod, Paris (1981).
- [11] J.M. Bismut, *Conjugate convex functions in optimal stochastic control*, J. Math. Anal. Appl. **44** (1973), 384–404.
- [12] J.M. Bismut, *Contrôle des système linéaire quadratiques: Applications de l’intégrale stochastique*, Sémin. Probab. XII, Lecture. Notes in Math., Vol. 649, Springer-Verlag (1978), 180–264.
- [13] B. Bouchard, *Stochastic target problem with*, Stochastic Process. Appl. (2002), 273–302.
- [14] B. Bouchard and N. Touzi, *Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations*, Stochastic Process. Appl. **111** (2004), 175–206.
- [15] K.A. Brakke, *The Motion of a Surface by Its Mean Curvature*, Princeton Univ. Press, Princeton, NJ (1978).
- [16] R. Buckdahn, P. Cardaliaguet and M. Quincampoix, *A representation formula for the mean curvature motion*, SIAM J. Math. Anal. **33** (4) (2002), 827–846.
- [17] R. Buckdahn and J. Ma, *Pathwise stochastic Taylor expansions and stochastic viscosity solutions for fully nonlinear stochastic PDEs*, Ann. Probab. **30** (2002), 1131–1171.
- [18] X. Cabre and L. Caffarelli, *Fully Nonlinear Elliptic Equations*, Amer. Math. Soc., Providence, RI (1995).

- [19] Y.-G. Chen, Y. Giga and S. Goto, *Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations*, J. Differential Geom. **33** (1991), 749–786.
- [20] P. Cheridito, H.M. Soner and N. Touzi, *The multi-dimensional super-replication problem under gamma constraints*, Ann. Inst. H. Poincaré Anal. Non Linéaire, **22** (5) (2005), 633–666.
- [21] P. Cheridito, H.M. Soner, N. Touzi and N. Victoir, *Second order backward stochastic differential equations and fully non-linear parabolic PDEs*, Comm. Pure and Applied Math. forthcoming.
- [22] D. Chevance, *Numerical Methods for Backward Stochastic Differential Equations*, Publ. Newton Inst., Cambridge University Press (1997).
- [23] M.G. Crandall, L.C. Evans and P.-L. Lions, *Some properties of viscosity solutions of Hamilton–Jacobi equations*, Trans. Amer. Math. Soc. **282** (1984), 487–502.
- [24] M.G. Crandall, H. Ishii and P.-L. Lions, *A user’s guide to viscosity solutions*, Bull. Amer. Math. Soc. **27** (1992), 1–67.
- [25] M.G. Crandall and P.-L. Lions, *Viscosity solutions of Hamilton–Jacobi equations*, Trans. Amer. Math. Soc. **277** (1984), 1–42.
- [26] F. Delarue and F. Menozzi, *A forward–backward stochastic algorithm for quasi-linear PDEs*, Preprint 932, University of Paris VI & VII (2004).
- [27] J. Douglas, J. Ma and P. Protter, *Numerical methods for forward-backward stochastic differential equations*, Ann. Appl. Probab. **6** (1996), 940–968.
- [28] E.B. Dynkin, *Superprocesses and partial differential equations*, Ann. Probab. **21** (3) (1993), 1185–1262.
- [29] E.B. Dynkin, *An Introduction to Branching Measure-Valued Processes*, CRM Monogr. Ser., Vol. 6, Amer. Math. Soc., Providence, RI (1994).
- [30] E.B. Dynkin, (1998), *A probabilistic approach to a nonlinear differential equation on a Riemannian manifold*, Teor. Veroyatnost. i Primenen. **42** (2) (1997), 336–341 (in Russian); English transl.: Theory Probab. Appl. no. 2, 289–294.
- [31] E.B. Dynkin, *Superdiffusions and Positive Solutions of Nonlinear Partial Differential Equations*, Univ. Lecture Ser., Vol. 34, Amer. Math. Soc., Providence, RI (2004).
- [32] N. El Karoui, S. Peng and M.C. Quenez, *Backward stochastic differential equations in finance*, Math. Finance **7** (1) (1997), 1–71.
- [33] L.C. Evans and L. Spruck, *Motion of level sets by mean curvature*, J. Differential Geom. **33** (1991), 635–681.
- [34] R.P. Feynman, *Space–time approach to nonrelativistic quantum mechanics*, Rev. Modern Phys. **20** (1948), 367–387.
- [35] W.H. Fleming and H.M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Appl. Math., Vol. 25, Springer-Verlag, New York (1993).
- [36] W.H. Fleming and P.E. Souganidis, *On the existence of value function of two-player, zero-sum stochastic differential games*, Indiana Univ. Math. J. **38** (1989), 293–314.
- [37] M. Freidlin, *Markov Processes and Differential Equations: Asymptotic Problems*, Lectures Math. ETH Zürich, Birkhäuser, Basel (1996).
- [38] M.I. Freidlin and A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, New York (1984).
- [39] T. Fujiwara and H. Kunita, *Stochastic differential equations of jump type and Levy processes in diffeomorphism groups*, J. Math. Kyoto Univ. **25** (1989), 71–106.
- [40] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, 2nd Edition, North-Holland (1989).
- [41] H. Ishii, *Uniqueness of unbounded viscosity solutions of Hamilton–Jacobi equations*, Indiana U. Math. J. **33** (1984), 721–748.
- [42] M. Kac, *On distributions of certain Wiener functional*, Trans. Amer. Math. Soc. (1949), 1–13.
- [43] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd Edition, Springer-Verlag (1991).
- [44] N.V. Krylov, *Nonlinear elliptic and Parabolic Partial Differential Equations of Second Order*, Math. Appl., Reider (1987).
- [45] N.V. Krylov, *Controlled Diffusion Processes*, Springer-Verlag, New York (1980).
- [46] O.A. Ladyzenskaya, V.A. Solonnikov and N.N. Uraltseva, *Linear and Quasilinear Equations of Parabolic Type*, Amer. Math. Soc., Providence, RI (1967).
- [47] P.-L. Lions, *Neumann type boundary conditions for Hamilton–Jacobi equations*, Duke J. Math. **52** (1985), 793–820.

- [48] P.L. Lions and H. Regnier, *Calcul du prix et des sensibilités d'une option américaine par une méthode de Monte Carlo*, Preprint (2001).
- [49] P.-L. Lions and P.E. Souganidis, *Fully nonlinear stochastic partial differential equations: Non-smooth equations and applications*, CRAS **327** (1998), 735–741.
- [50] P.-L. Lions and P.E. Souganidis, *Uniqueness of weak solutions of fully nonlinear stochastic partial differential equations*, CRAS **331** (2000), 783–790.
- [51] P.-L. Lions and A.S. Sznitman, *Stochastic differential equations with reflecting boundary conditions*, Comm. Pure Appl. Math. **37** (1984), 511–537.
- [52] J. Ma, P. Protter, J. San Martín and S. Torres, *Numerical methods for backward stochastic differential equations*, Ann. Appl. Probab. **12** (2002), 302–316.
- [53] J. Ma, P. Protter and J. Yong, *Solving backward stochastic differential equations explicitly – A four step scheme*, Probab. Theory Related Fields **98** (1994), 339–359.
- [54] B. Oksendal and A. Sulem, *Applied Stochastic Control of Jump Diffusions*, Springer-Verlag, New York (2005).
- [55] E. Pardoux and S. Peng, *Adapted solution of a backward stochastic differential equation*, Systems Control Lett. **14** (1990), 55–61.
- [56] E. Pardoux and S. Peng, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, Lecture Notes in Control and Inform. Sci., Vol. 176, Springer-Verlag (1992), 200–217.
- [57] E. Pardoux and S. Peng, *Backward doubly stochastic differential equations and systems of quasilinear parabolic SPDEs*, Probab. Theory Related Fields **98** (1994), 209–227.
- [58] S. Peng, *A general stochastic maximum principle for optimal control problems*, SIAM J. Control Optim. **28** (1990), 966–979.
- [59] S. Peng, *Probabilistic interpretation for systems of quasilinear parabolic partial differential equations*, Stochastics **37** (1991), 61–74.
- [60] A. Sayah, *Equations d'Hamilton–Jacobi du premier ordre avec termes intégrés différentiels*. Parties I and II, Comm. Partial Differential Equations **16** (1991), 1057–1093.
- [61] A.V. Skorokhod, *Studies in the Theory of Random Processes*, Dover, New York (1980); 1st Edition (1965).
- [62] H.M. Soner, *Jump Markov processes and viscosity solutions*, IMA Vol. Math. Appl., Vol. 10, Springer-Verlag, New York (1986), 501–511.
- [63] H.M. Soner, *Motion of a set by the curvature of its boundary*, J. Differential Equations **101** (1993), 313–372.
- [64] H.M. Soner and N. Touzi, *Stochastic target problems, dynamic programming and viscosity solutions*, SIAM J. Control Optim. **41** (2002), 404–424.
- [65] H.M. Soner and N. Touzi, *Dynamic programming for stochastic target problems and geometric flows*, J. Eur. Math. Soc. **4** (2002), 201–236.
- [66] H.M. Soner and N. Touzi, *A stochastic representation for mean curvature type geometric flows*, Ann. Probab. **31** (2002), 1145–1165.
- [67] H.M. Soner and N. Touzi, *Level set characterization of stochastic target problems*, Comm. Partial Differential Equation **27** (9/10) (2002), 2031–2053.
- [68] D. Talay, *Probabilistic numerical methods for partial differential equations: Elements of analysis*, Probabilistic Models for Nonlinear Partial Differential Equations, D. Talay and L. Tubaro, eds, Lecture Notes in Math., Vol. 1627 (1996), 48–196.

CHAPTER 7

Controllability and Observability of Partial Differential Equations: Some Results and Open Problems

Enrique Zuazua

Departamento de Matemáticas, Universidad Autónoma, 28049 Madrid, Spain

E-mail: enrique.zuazua@uam.es

Contents

1. Introduction	529
2. Preliminaries on finite-dimensional systems	531
2.1. Problem formulation	531
2.2. Controllability $\equiv$ observability	532
2.3. Bang-bang controls	538
3. Controllability of the linear wave equation	540
3.1. Statement of the problem	540
3.2. Exact controllability	543
3.3. Ingham inequalities and Fourier series techniques	550
3.4. Approximate controllability	553
3.5. Quasibang-bang controls	556
3.6. Stabilization	559
4. The heat equation	562
4.1. Problem formulation	562
4.2. Approximate controllability	564
4.3. Null controllability	568
4.4. Parabolic equations of fractional order	572
4.5. Carleman inequalities for heat equations with potentials	574
4.6. Bang-bang controls	574
4.7. Discussion and comparison	577
5. Sharp observability estimates for the linear heat equation	578
5.1. Sharp estimates	578
5.2. Optimality	583
6. Parabolic equations with low regularity coefficients	587
6.1. Elliptic equations	588
6.2. Parabolic equations	592

HANDBOOK OF DIFFERENTIAL EQUATIONS

Evolutionary Equations, volume 3

Edited by C.M. Dafermos and E. Feireisl

© 2007 Elsevier B.V. All rights reserved

7. Fluid–structure interaction models	596
7.1. Problem formulation	596
7.2. The model	600
7.3. Decay properties	604
8. Some open problems	606
Acknowledgements	613
References	614

Abstract

In this chapter we present some of the recent progresses done on the problem of controllability of partial differential equations (PDE). Control problems for PDE arise in many different contexts and ways. A prototypical problem is that of controllability. Roughly speaking it consists in analyzing whether the solution of the PDE can be driven to a given final target by means of a control applied on the boundary or on a subdomain of the domain in which the equation evolves. In an appropriate functional setting this problem is equivalent to that of observability which concerns the possibility of recovering full estimates on the solutions of the uncontrolled adjoint system in terms of partial measurements done on the control region. Observability/controllability properties depend in a very sensitive way on the class of PDE under consideration. In particular, heat and wave equations behave in a significantly different way, because of their different behavior with respect to time reversal. In this paper we first recall the known basic controllability properties of the wave and heat equations emphasizing how their different nature affects their main controllability properties. We also recall the main tools to analyze these problems: the so-called Hilbert uniqueness method (HUM), multipliers, microlocal analysis and Carleman inequalities. We then discuss some more recent developments concerning equations with low regularity coefficients, equations with potentials, bang-bang controls, etc. We also analyze the way control and observability properties depend on the norm and regularity of these coefficients, a problem which is also relevant when addressing nonlinear models. We then present some recent results on coupled models of wave–heat equations arising in fluid–structure interaction. We also present some open problems and future directions of research.

1. Introduction

In this chapter we address some topics related to the controllability of partial differential equations (PDE) which, in the context of Control Theory, are also often referred to as distributed parameter systems (DPS).

The controllability problem may be formulated roughly as follows. Consider an evolution system (either described in terms of partial or ordinary differential equations) on which we are allowed to act by means of a suitable choice of the control (the right-hand side of the system, the boundary conditions, etc.). Given a time interval $0 < t < T$, and initial and final states, the goal is to determine whether there exists a control driving the given initial data to the given final ones in time T .

This is a classical problem in Control Theory and there is a large literature on it. We refer, for instance, to the book of Lee and Markus [111] for an introduction to the topic in the context of finite-dimensional systems described in terms of ordinary differential equations (ODE). We also refer to the survey paper by Russell [150] and to the *SIAM Review* article and book by J.-L. Lions [112] and [113] for an introduction to the case of systems modeled by means of PDE.

There has been a very intensive research in this area in the last three decades and it would be impossible in this chapter to report on the main progresses that have been made. For this reason we have chosen a number of specific topics to present some recent results. Our goal is to exhibit the variety and depth of the problems arising in this field and some of the mathematical tools that have been used and developed to deal with them. Of course, the list of topics we have chosen is limited and it is not intended to represent the whole field. We hope however that, through this chapter, the reader will become familiar with some of the main research topics in this area. We have also included a long (but still incomplete) list of references for those readers interested in pursuing the study in this field and also a list of open problems for future research. As we shall see, many of them are closely related to other subtle questions of the theory of PDE, as unique continuation, asymptotic behavior of coupled systems, spectral properties, etc.

Even in the specific context of PDE, in order to address controllability problems in a successful way, one has still to make further distinctions between linear and nonlinear systems, time-reversible and time-irreversible ones, etc. In this chapter we mainly focus on linear problems and discuss both the wave and the heat equations, as the two main prototypes of reversible and irreversible models.

The techniques we present for the wave equation apply, essentially, to other models like, for instance, Schrödinger and plate equations. Combining them with fixed point arguments, these results may be extended to some semilinear models too. But, other relevant issues like, for instance, the bilinear control of Schrödinger equations need important further developments and different techniques that we shall not develop in this article. At this respect we refer to the recent work by Beauchard [8] (see also [9] for a global version of the same result and the references therein) where this problem is solved by a combination of several tools including Coron's return method ([32] and [35]) and Nash–Moser's iteration.

On the other hand, the techniques we shall present on the use of Carleman inequalities and variational methods for the control of the heat equation, strongly inspired in the works by Fursikov and Imanuvilov [76], can be extended to a wider class of parabolic problems.

In particular, with further important technical developments, this allows proving the local null controllability of Navier–Stokes equations. We refer to [68] for the latest results on this problem and to [67] for a survey on that topic. The Euler equations are also well known to be controllable (see [33] and [80]). However, because of the hyperbolic nature of the problem, this time Carleman inequalities may not be applied but rather the return method needs to be used.

Taking into account that the existing theory is able to cover quite successfully both hyperbolic and parabolic models or, in other words, structures and fluids, it is natural to address the important issue of fluid–structure interaction. Recently important progresses have been made also in this context too. First, existence results are available for a number of models in which the structure is considered to be a rigid body [151] or a flexible one [13, 14, 39]. Part of this chapter will be devoted to report on these results. But we shall mainly focus on a simplified linearized model in which the wave and heat equation are coupled through a fixed interface. We shall mainly discuss the problem of the asymptotic behavior of solutions. The techniques developed for the controllability of the wave equation will play a key role when doing that. As we shall see, some of the dynamical properties of the system we shall describe could seem unexpected. For instance, the damping effect that the heat equation introduces on the wave solutions is too strong and overdamping occurs and the decay rate fails to be exponentially uniform. The problem of controllability is by now only well understood in one space dimension. There is still to be done in this field to address controllability in several space dimensions and then for covering the nonlinear free boundary problems. One of the very few existing results on the subject is that in [15] that guarantees the local controllability of the Navier–Stokes equations, coupled with moving rigid bodies.

In this chapter we do not address the issue of numerical approximation of controls. This is, of course, a very important topic for the implementation of the control theoretical results in practical applications. We refer to [180] for a recent survey article in this issue (see also [178]) and [179] for a discussion in connection with optimal control problems.

As we said above, the choice of the topics in this article is necessarily limited. The interested reader may complement these notes with the survey articles [173, 176] for the controllability of PDE, and [27] for the controllability and homogenization. We also refer to the notes [127] for an introduction to some of the most elementary tools in the controllability of PDE. The notes [127] are in fact published in a collective book which contains interesting survey and introductory papers in Control Theory. The article [168] contains a discussion of the state of the art on the controllability of semilinear wave equations, published in a collective work on unsolved problems in Control Theory that might be of interest for researchers in this area. However, our bibliography is not complete. There are, for instance, other books related to this and other closely related topics as, for instance, [57, 59] and [105] and they contain many other bibliographical complements.

The content of this chapter is as follows. In Section 2 we make a brief introduction to the topic in the context of linear finite-dimensional systems. Sections 3 and 4 are devoted to describe the main issues related to the controllability of the linear wave and heat equations, respectively, and the basic known results. We also discuss in detail the existence of bang-bang controls. In Section 5 we discuss the optimality of the known observability results for heat equations with potentials. We show that in the context of multidimensional

parabolic systems the existing observability estimates are indeed sharp in what concerns the dependence on the L^∞ -norm of the potential. In Section 6 we present some simple but new results on the observability of the heat equation with low regularity variable coefficients on the principal part, a topic which is full of interesting and difficult open problems. In Section 7 we discuss some models coupling heat and wave equations along a fixed interface, which may be viewed as a simplified and linearized version of more realistic models of fluid–structure interaction. We end up with a section devoted to present some open problems and future directions of research.

2. Preliminaries on finite-dimensional systems

2.1. Problem formulation

PDE can be viewed as infinite-dimensional versions of linear systems of ordinary differential equations (ODE). ODE generate finite-dimensional dynamical systems, while PDE correspond to infinite-dimensional ones. The fact that PDE are an infinite-dimensional version of finite-dimensional ODE can be justified and is relevant in various different contexts. First, that is the case in Mechanics. While PDE are the common models for Continuum Mechanics, ODE arise in classical Mechanics, where the continuous aspect of the media under consideration is not taken into account. The same can be said in the context of Numerical Analysis. Numerical approximation schemes for PDE and, more precisely, those that are semidiscrete (discrete in space and continuous in time) yield finite-dimensional systems of ODE. This is particularly relevant in the context of control where, when passing to the limit from finite to infinite dimensions, unwanted and unexpected pathologies may arise (see [180]).

It is, therefore, convenient to first have a quick look to the problems under consideration in the finite-dimensional context. This will be useful when dealing with singular limits from finite- to infinite-dimensional systems and, in particular, when addressing numerical approximation issues. But it will also be useful to better understand the problems and techniques we shall use in the context of PDE, where things are necessarily technically more involved and complex due to the much richer structure associated to the continuous character of the media under consideration and the needed Functional Analytical tools.

There is by now an extensive literature on the control of finite-dimensional systems and the problem is completely understood for linear ones [111,155]. Here we shall only present briefly the problems and techniques we shall later employ in the context of PDE.

Consider the finite-dimensional system of dimension N

$$x' + Ax = Bv, \quad 0 \leq t \leq T, \quad x(0) = x^0, \quad (2.1)$$

where $x = x(t)$ is the N -dimensional *state* and $v = v(t)$ is the M -dimensional *control*, with $M \leq N$. By “ $'$ ” we denote differentiation with respect to time t .

Here A is an $N \times N$ matrix with constant real coefficients and B is an $N \times M$ matrix. The matrix A determines the dynamics of the system and the matrix B models the way M controls act on it.

In practice, it is desirable to control the N components of the system with a low number of controls and the best would be to do it by a single one, in which case $M = 1$. As we shall see, this is possible provided B , the control operator, is chosen appropriately with respect to the matrix A governing the dynamics of the system.¹

System (2.1) is said to be *controllable* in time T when every initial datum $x^0 \in \mathbb{R}^N$ can be driven to any final datum $x^1 \in \mathbb{R}^N$ in time T by a suitable control $v \in (L^2(0, T))^M$, i.e., the following final condition is satisfied

$$x(T) = x^1. \quad (2.2)$$

In other words the system is said to be controllable in time T when the set of reachable states,

$$R(T; x^0) = \{x(T): v \in (L^2(0, T))^M\},$$

covers the whole $\mathbb{R}^N$ and this for all $x^0 \in \mathbb{R}^N$. When this property holds, the system is said to be *exactly controllable*. Here “exactly” refers to the fact that the target (2.2) is achieved completely. This final condition can be relaxed in different ways leading to various weaker notions of controllability. However, as we shall see, since we are in finite dimensions, these apparently weaker notions often coincide with the exact controllability one. For instance, the system is said to be *approximately controllable* when the set of reachable states is dense in $\mathbb{R}^N$. But, in $\mathbb{R}^N$, the only close affine dense subspace is the whole space itself. Thus, approximate and exact controllability are equivalent notions.

But let us analyze the problem of exact controllability.

There is a necessary and sufficient condition for (exact) controllability which is of purely algebraic nature. It is the so-called *Kalman condition*: System (2.1) is controllable in some time $T > 0$ iff

$$\text{rank}[B, AB, \dots, A^{N-1}B] = N. \quad (2.3)$$

Moreover, when this holds, the system is controllable for all time $T > 0$.

Note that the matrix $[B, AB, \dots, A^{N-1}B]$ has to be considered as a line of blocks of $N \times MN$ elements.

There is a direct proof of this result which uses the representation of solutions of (2.1) by means of the variation of constants formula. However, for addressing PDE models it is more convenient to use an alternative method which consists in transforming the control problem into a problem of observability for the adjoint system, since the later one can be solved by a combination of the existing methods to obtain a priori estimates on solutions of ODE and PDE.

2.2. Controllability $\equiv$ observability

Let us introduce the problem of observability.

¹This being possible for appropriate choices of the control operator B allows us to be optimistic when addressing PDE models, in which case the state variable is infinite-dimensional.

Consider the *adjoint system*

$$-\varphi' + A^*\varphi = 0, \quad 0 \leq t \leq T, \quad \varphi(T) = \varphi^0. \quad (2.4)$$

The following fundamental result establishes the equivalence between the controllability of system (2.1) and the observability property of the adjoint system (2.4).

THEOREM 2.1. *System (2.1) is controllable in time T if and only if the adjoint system (2.4) is observable in time T , i.e., if there exists a constant $C = C(T) > 0$ such that, for all solution φ of (2.4),*

$$|\varphi^0|^2 \leq C \int_0^T |B^*\varphi|^2 dt. \quad (2.5)$$

Both properties hold in all time T if and only if the Kalman rank condition (2.3) is satisfied.

SKETCH OF THE PROOF. We first prove that the observability inequality (2.5) for the adjoint system (2.4) implies the controllability of the state equation (2.1). Our proof provides a constructive method to build controls.

We proceed in several steps.

STEP 1 (Construction of controls as minimizers of a quadratic functional). Assume (2.5) holds and consider the quadratic functional $J: \mathbb{R}^N \rightarrow \mathbb{R}$

$$J(\varphi^0) = \frac{1}{2} \int_0^T |B^*\varphi(t)|^2 dt - \langle x^1, \varphi^0 \rangle + \langle x^0, \varphi(0) \rangle. \quad (2.6)$$

If $\hat{\varphi}^0$ is a minimizer for J , $DJ(\hat{\varphi}^0) = 0$, and the control

$$v = B^*\hat{\varphi}, \quad (2.7)$$

where $\hat{\varphi}$ is the solution of (2.4) with that datum $\hat{\varphi}^0$ at time $t = T$, is such that the solution x of (2.1) satisfies the control requirement $x(T) = x^1$.

Indeed, for all $\psi^0, \varphi^0 \in \mathbb{R}^N$,

$$\langle DJ(\psi^0), \varphi^0 \rangle = \int_0^T B^*\psi(t) \cdot B^*\varphi(t) dt - \langle x^1, \varphi^0 \rangle + \langle x^0, \varphi(0) \rangle.$$

Thus, $DJ(\hat{\varphi}^0) = 0$ if and only if

$$\int_0^T B^*\hat{\varphi}(t) \cdot B^*\varphi(t) dt - \langle x^1, \varphi^0 \rangle + \langle x^0, \varphi(0) \rangle = 0$$

for all $\varphi^0 \in \mathbb{R}^N$. In other words,

$$\int_0^T BB^*\hat{\varphi}(t) \cdot \varphi(t) dt - \langle x^1, \varphi^0 \rangle + \langle x^0, \varphi(0) \rangle = 0,$$

or

$$\int_0^T Bv \cdot \varphi(t) dt - \langle x^1, \varphi^0 \rangle + \langle x^0, \varphi(0) \rangle = 0 \quad (2.8)$$

if v is chosen according to (2.7).

Here and in the sequel we denote by “ $\cdot$ ” or $\langle \cdot, \cdot \rangle$ the scalar product in the Euclidean space (both in $\mathbb{R}^N$ and $\mathbb{R}^M$).

We claim that (2.8) is equivalent to the fact that the control v as above drives the solution x of (2.1) from x^0 to x^1 . Indeed, multiplying the state equation (2.1) by any solution φ of the adjoint system (2.4), we get

$$\int_0^T (x' + Ax) \cdot \varphi dt = \int_0^T Bv \cdot \varphi dt. \quad (2.9)$$

On the other hand,

$$\begin{aligned} \int_0^T (x' + Ax) \cdot \varphi dt &= \int_0^T x \cdot (-\varphi' + A^* \varphi) dt + \langle x, \varphi \rangle|_0^T \\ &= \langle x(T), \varphi^0 \rangle - \langle x^0, \varphi(0) \rangle. \end{aligned} \quad (2.10)$$

Combining (2.8) and (2.10) we deduce that

$$\langle x(T) - x^1, \varphi^0 \rangle = 0,$$

for all $\varphi^0 \in \mathbb{R}^N$. This is equivalent to the final condition (2.2) imposed to the control problem.

Thus, to solve the control problem it is sufficient to prove that the functional J in (2.6) achieves a minimizer. To do that, we apply the direct method of the Calculus of Variations (DMCV). The functional J being continuous, quadratic and convex, and defined in the finite-dimensional Euclidean space, it is sufficient to prove its coercivity, i.e.,

$$\lim_{\|\varphi^0\| \rightarrow \infty} J(\varphi^0) = \infty. \quad (2.11)$$

This property holds if and only if the observability inequality is satisfied. Indeed, when (2.5) holds the following variant holds as well, with possibly a different constant $C > 0$,

$$|\varphi^0|^2 + |\varphi(0)|^2 \leq C \int_0^T |B^* \varphi|^2 dt. \quad (2.12)$$

In fact, both inequalities (2.5) and (2.12) are equivalent. This is so since $\varphi(t) = e^{A^*(t-T)} \varphi_0$ and the operator $e^{A^*(t-T)}$ is bounded and invertible.

In view of (2.12) the coercivity of J follows. This implies the existence of the minimizer for J and therefore that of the control we are looking for.

STEP 2 (Equivalence between the observability inequality (2.12) and the Kalman condition). In the previous step we have shown that the observability inequality (2.12) implies the existence of the control. In this second step we show that the observability inequality is equivalent to the Kalman condition.

Since we are in finite dimension and all norms are equivalent, (2.12) is equivalent to the following uniqueness property:

$$\text{Does the fact that } B^* \varphi \text{ vanish for all } 0 \leq t \leq T \text{ imply that } \varphi \equiv 0. \quad (2.13)$$

Taking into account that solutions φ of the adjoint system are analytic in time, $B^* \varphi$ vanishes if and only if all the derivatives of $B^* \varphi$ of any order vanish at time $t = T$. Since $\varphi = e^{A^*(t-T)} \varphi^0$ this is equivalent to the fact that $B^* [A^*]^k \varphi^0 \equiv 0$ for all $k \geq 0$. But, according to the Cayley–Hamilton theorem, this holds if and only if it is satisfied for all $k = 0, \dots, N-1$. Therefore $B^* \varphi \equiv 0$ is equivalent to $[B^*, B^* A^*, \dots, B^* [A^*]^{N-1}] \varphi^0 = 0$. But, the latter, when

$$\text{rank}[B^*, B^* A^*, \dots, B^* [A^*]^{N-1}] = N,$$

is equivalent to the fact that $\varphi^0 = 0$ or $\varphi \equiv 0$. Obviously, this rank condition is equivalent to the Kalman one (2.3). Here, the matrix $[B^*, B^* A^*, \dots, B^* [A^*]^{N-1}]$ has to be considered as a column of blocks with $MN \times N$ elements.

This concludes the proof of the fact that observability implies controllability. Let us now prove the reverse assertion, i.e., that controllability implies observability.

Let us assume that the state equation is controllable. We choose $x^1 = 0$. Then, for all $x^0 \in \mathbb{R}^N$ there exists a control $v \in (L^2(0, T))^M$ such that the solution of (2.1) satisfies $x(T) = 0$. The control is not unique thus it is convenient to choose the one of minimal norm. By the closed graph theorem, we deduce that there exists a constant $C > 0$ (that, in particular, depends on the control time T) such that

$$\|v\|_{(L^2(0, T))^M} \leq C |x^0|. \quad (2.14)$$

Then, multiplying the state equation (2.1) by any solution of the adjoint equation φ and taking into account that $x(T) = 0$ for the control v we have chosen, we deduce that

$$-\langle x^0, \varphi(0) \rangle = \int_0^T v \cdot B^* \varphi \, dt.$$

Combining this identity with (2.14) we deduce that

$$|\langle x^0, \varphi(0) \rangle| \leq C |x^0| \|B^* \varphi\|_{(L^2(0, T))^M}$$

for all $x^0 \in \mathbb{R}^N$, which is equivalent to

$$|\varphi(0)| \leq C \|B^* \varphi\|_{(L^2(0, T))^M}. \quad (2.15)$$

This estimate (2.15) is equivalent to the observability inequalities (2.5) and/or (2.12). This is so, once more, because of the continuity of the mapping $\varphi(0) \rightarrow \varphi^0$. $\square$

REMARK 2.1. The property of observability of the adjoint system (2.4) is equivalent to the inequality (2.5) because of the linear character of the system. In general, the problem of observability can be formulated as that of determining uniquely the adjoint state everywhere in terms of partial measurements.

We emphasize that, in the finite-dimensional context under consideration, the observability inequality (2.5) is completely equivalent to (2.12) and/or (2.15). In other words, it is totally equivalent to formulate the problem of estimating the initial or final data of the adjoint system. This is so because the mapping $\varphi^0 \rightarrow \varphi(0)$ is continuous, and has continuous inverse. This is no longer necessarily true for infinite-dimensional systems. This fails, in particular, for time-irreversible equations as the heat equation.

There is another major difference with infinite-dimensional systems written in terms of PDEs. Namely, the uniqueness property (2.13) may hold but this does not necessarily imply an observability inequality (2.5) to be true in the desired energy space. This is due to the fact that, in infinite-dimensional Banach spaces, all norms are not necessarily equivalent. In other words, in infinite dimension a strict subspace may be dense, and this never occurs in finite dimension.

REMARK 2.2. This proof of controllability provides a constructive method to build the control: minimizing the functional J . But it also yields explicit bounds on the controls. Indeed, since the functional $J \leq 0$ at the minimizer, and in view of the observability inequality (2.12), it follows that

$$\|v\| \leq 2\sqrt{C} [|x^0|^2 + |x^1|^2]^{1/2}, \quad (2.16)$$

C being the same constant as in (2.12). Therefore, we see that the observability constant is, up to a multiplicative factor, the norm of the control map associating to the initial and final data of the state equation (x^0, x^1) the control of minimal norm v . Actually, a more careful analysis indicates that the norm of the control can be bounded above in terms of the norm of $e^{AT}x^0 - x^1$ which measures the distance between the target x^1 and the final state $e^{AT}x^0$ that the uncontrolled dynamics would reach without implementing any control.

Our proof above shows that the reverse is also true. In other words, the norm of the control map that associates the control v to each pair of initial/final data (x^0, x^1) , also provides an explicit observability constant.

REMARK 2.3. Furthermore, the approach above has also the interesting property of providing systematically the control of minimal $L^2(0, T)$ -norm within the class of admissible ones. Indeed, given T , an initial datum and a final one, if the system is controllable, there are infinitely many controls driving the trajectory from the initial datum to the final target. To see this it is sufficient to argue as follows. In the first half of the time interval $[0, T/2]$ we can choose any function as controller. This drives the system to a new state, say, at time $T/2$. The system being controllable, it is controllable in the second half of the time interval $[T/2, T]$. This allows applying the variational approach above to obtain the control driving

the system from its value at time $t = T/2$ to the final state x^1 at time T in that second interval. The superposition of these two controls provides an admissible control which has an arbitrary shape in the first interval $[0, T/2]$. This suffices to see that the set of admissible controls contains an infinite number of elements.

As we said above, the variational approach we have described provides the control of minimal $L^2(0, T)$ -norm. Indeed, assume, to simplify the presentation, that $x^1 = 0$. Let u be an arbitrary control and v the control we have constructed by the variational approach. Multiplying by φ in the state equation and integrating by parts with respect to time, we deduce that both controls satisfy

$$\int_0^T \langle u, B^* \varphi \rangle dt = \int_0^T \langle v, B^* \varphi \rangle dt = -\langle x^0, \varphi(0) \rangle$$

for any solution φ of the adjoint system. In particular, by taking $\hat{\varphi}$, the solution of the adjoint system corresponding to the minimizer of J and that determines the control v (i.e., $v = B^* \hat{\varphi}$), it follows that

$$\int_0^T \langle u, B^* \hat{\varphi} \rangle dt = \int_0^T \langle v, B^* \hat{\varphi} \rangle dt = \int_0^T |v|^2 dt = -\langle x^0, \hat{\varphi}(0) \rangle.$$

Thus,

$$\begin{aligned} \|v\|_{L^2(0,T)}^2 &\leq \left| \int_0^T \langle u, B^* \hat{\varphi} \rangle dt \right| \leq \|u\|_{L^2(0,T)} \|B^* \hat{\varphi}\|_{L^2(0,T)} = \|u\|_{L^2(0,T)} \|v\|_{L^2(0,T)}, \end{aligned}$$

which implies that $\|v\|_{L^2(0,T)} \leq \|u\|_{L^2(0,T)}$. This completes the proof of the minimality of the control we have built by the variational approach.

REMARK 2.4. It is important to note that, in this finite-dimensional context, the time T of controllability/observability plays no role. Of course this is true, in particular, because the system under consideration is autonomous. In particular, whether a system is controllable (or its adjoint observable) is independent of the time T of control since these properties only depend on the algebraic Kalman condition. Note that the situation may be totally different for PDE. In particular, as we shall see, in the context of the wave equation, due to the finite velocity of propagation, the time needed to control/observe waves from the boundary needs to be large enough, of the order of the ratio between size of the domain and velocity of propagation.

REMARK 2.5. The set of controllable pairs (A, B) is open and dense. Indeed:

- If (A, B) is controllable there exists $\varepsilon > 0$ sufficiently small such that any (A^0, B^0) with $|A^0 - A| < \varepsilon$, $|B^0 - B| < \varepsilon$ is also controllable. This is a consequence of the Kalman rank condition and of the fact that the determinant of a matrix depends continuously on its entries.

This shows the robustness of the controllability property under (small) perturbations of the system.

- On the other hand, if (A, B) is not controllable, for any $\varepsilon > 0$, there exists (A^0, B^0) with $|A - A^0| < \varepsilon$ and $|B - B^0| < \varepsilon$ such that (A^0, B^0) is controllable. This is a consequence of the fact that the determinant of an $N \times N$ matrix depends analytically on its entries and cannot vanish in a ball of $\mathbb{R}^n$.

2.3. Bang-bang controls

In the previous section we have proved the equivalence of the controllability property of the state equation and the observability property for the adjoint. This has been done in the $L^2(0, T)$ -setting and we have developed a variational method allowing to obtain the control of minimal $L^2(0, T)$ -norm, which turns out to be C^∞ smooth and even analytic in time, in view of its structure (2.7).

Smooth controllers are however difficult to implement in practice because of its continuous and subtle change in shape and intensity. In the opposite extreme we may think on bang-bang controls which are piecewise constant and discontinuous but easier to implement since they consist simply in switching from a constant value to another. Once the size of the bang-bang control is determined, it is completely identified by the location of the switching times.

The goal of this section is to show that, with the ideas we have developed before and some minor changes, one can show that, whenever the system is controllable, bang-bang controls exist, and to give a variational procedure to compute them.

To simplify the presentation, without loss of generality, we suppose that $x^1 \equiv 0$.

In order to build bang-bang controls, it is convenient to consider the quadratic functional

$$J_{\text{bb}}(\varphi^0) = \frac{1}{2} \left[\int_0^T |B^* \varphi| dt \right]^2 + \langle x^0, \varphi(0) \rangle, \quad (2.17)$$

where φ is the solution of the adjoint system (2.4) with initial data φ^0 at time $t = T$.

It is interesting to note that J_{bb} differs from J in the quadratic term. Indeed, in J we took the $L^2(0, T)$ -norm of $B^* \varphi$ while here we consider its $L^1(0, T)$ -norm.

The functional J_{bb} is continuous, convex and also coercive because the unique continuation property (2.13) holds. It follows that J_{bb} attains a minimum in some point $\hat{\varphi}^0 \in \mathbb{R}^N$. This can be easily seen using the direct method of the Calculus of Variations and taking into account that, in $\mathbb{R}^N$, all bounded sequences are relatively compact.

Note that for the coercivity of J_{bb} to hold one needs the following L^1 -version of the observability inequality (2.5):

$$|\varphi^0|^2 \leq C \left[\int_0^T |B^* \varphi| dt \right]^2. \quad (2.18)$$

This inequality holds immediately as a consequence of the unique continuation property (2.13) because we are in the finite-dimensional setting. However, in the infinite-dimensional setting things might be much more complex in the sense that the unique

continuation property does not imply any specific observability inequality automatically. This will be particularly relevant when analyzing wave-like equations.

On the other hand, it is easy to see that

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \left[\left(\int_0^T |f + hg| dt \right)^2 - \left(\int_0^T |f| dt \right)^2 \right] \\ = 2 \int_0^T |f| dt \int_0^T \operatorname{sgn}(f(t))g(t) dt \end{aligned} \quad (2.19)$$

if the Lebesgue measure of the set $\{t \in (0, T): f(t) = 0\}$ vanishes.

Here and in the sequel the sign function “sgn” is defined as a multivalued function in the following way

$$\operatorname{sgn}(s) = \begin{cases} 1 & \text{when } s > 0, \\ [-1, 1] & \text{when } s = 0, \\ -1 & \text{when } s < 0. \end{cases}$$

Remark that in the previous limit there is no ambiguity in the definition of $\operatorname{sgn}(f(t))$ since the set of points $t \in [0, T]$ where $f = 0$ is assumed to be of zero Lebesgue measure and does not affect the value of the integral.

Identity (2.19) may be applied to the quadratic term of the functional J_{bb} since, taking into account that φ is the solution of the adjoint system (2.4), it is an analytic function and therefore, each of the components of $B^*\varphi$ changes sign finitely many times in the interval $[0, T]$ except when $\hat{\varphi}^0 = 0$. Rigorously speaking, this is true when each of the pairs (A, b_i) , b_i is being the column vectors of B , satisfies the Kalman rank condition. In view of this, the Euler–Lagrange equation associated with the critical points of the functional J_{bb} is as follows:

$$\int_0^T |B^*\hat{\varphi}| dt \int_0^T \operatorname{sgn}(B^*\hat{\varphi}) B^*\psi(t) dt + \langle x^0, \psi(0) \rangle = 0$$

for all $\psi^0 \in \mathbb{R}^N$, where ψ is the solution of the adjoint system (2.4) with initial data ψ^0 . When applied to a vector, $\operatorname{sgn}(\cdot)$ is defined componentwise as before.

Consequently, the control we are looking for is $v = \int_0^T |B^*\hat{\varphi}| dt \operatorname{sgn}(B^*\hat{\varphi})$, where $\hat{\varphi}$ is the solution of (2.4) with initial data $\hat{\varphi}^0$, the minimizer of J_{bb} .

Note that when $M = 1$, i.e., when the control u is a scalar function, it is of bang-bang form. Indeed, v takes only two values $\pm \int_0^T |B^*\hat{\varphi}| dt$. The control switches from one to the other one when the function $B^*\hat{\varphi}$ changes sign. This happens finitely many times. When $M > 1$, the control v is a vector valued bang-bang function in the sense that each component is of bang-bang form. Note however that each component of v may change sign in different times, depending on the changes of sign of the corresponding component of $B^*\hat{\varphi}$.

REMARK 2.6. Other types of controls can be obtained by considering functionals of the form

$$J_p(\varphi^0) = \frac{1}{2} \left(\int_0^T |B^* \varphi|^p dt \right)^{2/p} + \langle x^0, \varphi^0 \rangle$$

with $1 < p < \infty$. The corresponding controls are

$$v = \left(\int_0^T |B^* \hat{\varphi}|^p dt \right)^{(2-p)/p} |B^* \hat{\varphi}|^{p-2} B^* \hat{\varphi},$$

where $\hat{\varphi}$ is the solution of (2.4) with initial datum $\hat{\varphi}^0$, the minimizer of J_p .

It can be shown that, as expected, the controls obtained by minimizing these functionals give, in the limit when $p \rightarrow 1$, a bang-bang control.

In the previous section we have seen that the control obtained by minimizing the functional J is of minimal $L^2(0, T)$ -norm. We claim that the control obtained by minimizing the functional J_{bb} is of minimal $L^\infty(0, T)$ -norm. Indeed, let u be any control in $L^\infty(0, T)$ and v be the one obtained by minimizing J_{bb} . Once more we have

$$\int_0^T u \cdot B^* \hat{\varphi} dt = \int_0^T v \cdot B^* \hat{\varphi} dt = -\langle x^0, \hat{\varphi}(0) \rangle.$$

In view of the definition of v , it follows that

$$\|v\|_{L^\infty(0, T)}^2 = \left(\int_0^T |B^* \hat{\varphi}| dt \right)^2 = \int_0^T u \cdot B^* \hat{\varphi} dt = \int_0^T v \cdot B^* \hat{\varphi} dt.$$

Hence,

$$\begin{aligned} \|v\|_{L^\infty(0, T)}^2 &= \int_0^T u \cdot B^* \hat{\varphi} dt \leq \|u\|_{L^\infty(0, T)} \int_0^T |B^* \hat{\varphi}| dt \\ &= \|u\|_{L^\infty(0, T)} \|v\|_{L^\infty(0, T)} \end{aligned}$$

and the proof finishes.

3. Controllability of the linear wave equation

3.1. Statement of the problem

Let Ω be a bounded domain of $\mathbb{R}^n$, $n \geq 1$, with boundary Γ of class C^2 . Let ω be an open and nonempty subset of Ω and $T > 0$.

Consider the linear controlled wave equation in the cylinder $Q = \Omega \times (0, T)$

$$\begin{cases} u_{tt} - \Delta u = f \mathbb{1}_\omega & \text{in } Q, \\ u = 0 & \text{on } \Sigma, \\ u(x, 0) = u^0(x), \quad u_t(x, 0) = u^1(x) & \text{in } \Omega. \end{cases} \quad (3.1)$$

In (3.1) Δ denotes the Laplacian, Σ represents the lateral boundary of the cylinder Q , i.e., $\Sigma = \Gamma \times (0, T)$, $\mathbb{1}_\omega$ is the characteristic function of the set ω , $u = u(x, t)$ is the state and $f = f(x, t)$ is the control variable. Since f is multiplied by $\mathbb{1}_\omega$ the action of the control is localized in ω .

When $(u^0, u^1) \in H_0^1(\Omega) \times L^2(\Omega)$ and $f \in L^2(Q)$ system (3.1) has a unique finite energy solution $u \in C([0, T]; H_0^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$.

The problem of *controllability* consists roughly in *describing the set of reachable final states*

$$R(T; (u^0, u^1)) = \{(u(T), u_t(T)) : f \in L^2(Q)\}.$$

It is the affine subspace of the final states that the solutions reach at time $t = T$, starting from the initial datum (u^0, u^1) , when the control f varies all over $L^2(Q)$. Note however that the action of the control is localized in ω . Thus, the controls may also be viewed to belong to $L^2(\omega \times (0, T))$.

One may distinguish the following notions of controllability:

- (a) *Approximate controllability.* System (3.1) is said to be approximately controllable in time T if the set of reachable states is dense in $H_0^1(\Omega) \times L^2(\Omega)$ for every $(u^0, u^1) \in H_0^1(\Omega) \times L^2(\Omega)$.
- (b) *Exact controllability.* System (3.1) is said to be exactly controllable at time T if

$$R(T; (u^0, u^1)) = H_0^1(\Omega) \times L^2(\Omega)$$

for all $(u^0, u^1) \in H_0^1(\Omega) \times L^2(\Omega)$.

- (c) *Null controllability.* System (3.1) is said to be null controllable at time T if

$$(0, 0) \in R(T; (u^0, u^1))$$

for all $(u^0, u^1) \in H_0^1(\Omega) \times L^2(\Omega)$.

REMARK 3.1.

(a) Since we are dealing with solutions of the wave equation, due to the finite speed of propagation, for any of these properties to hold the control time T has to be sufficiently large, the trivial case in which the control subdomain ω coincides with the whole domain Ω being excepted.

(b) Since system (3.1) is linear and reversible in time null and exact controllability are equivalent notions, as in the finite-dimensional case of the previous section. As we shall see, the situation is completely different in the case of the heat equation.

(c) Clearly, every exactly controllable system is approximately controllable too. However, system (3.1) may be approximately but not exactly controllable. Obviously, this does not happen in the context of finite-dimensional systems since exact and approximate controllability are equivalent notions. This is so because, in $\mathbb{R}^N$, the only affine dense subspace of $\mathbb{R}^N$ is $\mathbb{R}^N$ itself.

In those cases in which approximate controllability holds but exact controllability fails it is natural to study the *cost of approximate controllability*, or, in other words, the size of the control needed to reach an ε -neighborhood of a final state which is not exactly reachable. This problem was analyzed by Lebeau in [106] in the context of wave equations with analytic coefficients. Roughly speaking, when exact controllability fails, the cost of reaching a target which does not belong to the subspace of reachable data, increases exponentially as the distance ε to the target tends to zero. Later on a slightly weaker version of this result was given by Robbiano [147] in the context of wave equations with C^2 coefficients in the principal part, C^3 domains, and with lower order potentials, by means of Carleman inequalities.

(e) The controllability problem above may also be formulated in other function spaces in which the wave equation is well posed. For instance one can take initial and final data in $L^2(\Omega) \times H^{-1}(\Omega)$ and then the control in $L^2(0, T; H^{-1}(\omega))$ or, by the contrary, the initial data in $H^2 \cap H_0^1(\Omega) \times H_0^1(\Omega)$ and the control in $L^2(0, T; H_0^1(\omega))$. Similar results hold in all these cases. In these notes we have chosen to work in the classical context of finite-energy solutions of the wave equation to avoid unnecessary technicalities.

(f) Null controllability is a physically particularly interesting notion since the state $(0, 0)$ is an equilibrium for system (3.1). Once the system reaches the equilibrium at time $t = T$, we can stop controlling (by taking $f \equiv 0$ for $t \geq T$) and the system naturally stays in the equilibrium configuration for all $t \geq T$.

(g) Most of the literature on the controllability of the wave equation has been written on the framework of the *boundary control* problem. The control problems formulated above for system (3.1) are usually referred to as *internal controllability* problems since the control acts on the subset ω of Ω . Although the results are essentially the same in both cases, the boundary control problem is normally more complex from a technical point of view, because of the intrinsic difficulty of dealing with boundary traces and nonhomogeneous boundary value problems. The closer analogies arise when considering boundary control problems on one side and, on the other one, internal controls localized in ω , a neighborhood of the boundary of the domain Ω or part of it (see [112]).

In the context of boundary control the state equation reads

$$\begin{cases} u_{tt} - \Delta u = 0 & \text{in } Q, \\ u = v \mathbb{1}_{\Sigma_0} & \text{on } \Sigma, \\ u(x, 0) = u^0(x), u_t(x, 0) = u^1(x) & \text{in } \Omega, \end{cases} \quad (3.2)$$

where Σ_0 is the subset of the lateral boundary $\Sigma = \Gamma \times (0, T)$ where the control is applied. In most cases the subset of the boundary Σ_0 is taken to be cylindrical, i.e., $\Sigma_0 = \Gamma_0 \times (0, T)$ for a subset Γ_0 of $\partial\Omega$. But Σ_0 can be any nonempty relative open subset of the lateral boundary Σ . The most natural functional setting is that in which $v \in L^2(\Sigma_0)$ and $u \in C([0, T]; L^2(\Omega)) \cap C^1([0, T]; H^{-1}(\Omega))$. In this setting the formulation of approximate, exact and null control problems is basically the same, except for the fact that, from a

technical point of view, addressing them is more complex in this context of boundary control since one has to deal with fine trace results. In fact, proving that system (3.2) is well posed in $C([0, T]; L^2(\Omega)) \cap C^1([0, T]; H^{-1}(\Omega))$ with boundary data in $L^2(\Sigma_0)$ requires a quite subtle use of the method of transposition (see [112]). But the methods we shall develop further, based on the observability of the adjoint system, apply in this context too. The techniques we shall describe apply also to other boundary conditions. However, the analysis of well-posedness for the corresponding nonhomogeneous boundary value problems may present new difficulties (see [112]).

3.2. Exact controllability

In the previous section we have explained the equivalence between the controllability of the state equation and a suitable observability property for the adjoint system in the context of finite-dimensional systems. The same is true for PDE. But, as we mentioned above, the problem is much more complex in the context of PDE since we are dealing with infinite-dimensional dynamical systems and not all norms are equivalent in this setting.

In the context of PDE, the unique continuation property by itself, i.e., the PDE analogue of (2.13), does not suffice and one has to directly address the problem of observability paying special attention to the norms involved on the observability inequality.

In this case the adjoint system is as follows:

$$\begin{cases} \varphi_{tt} - \Delta \varphi = 0 & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x), \quad \varphi_t(x, T) = \varphi^1(x) & \text{in } \Omega. \end{cases} \quad (3.3)$$

As it was shown by Lions [113], using the so called HUM (Hilbert uniqueness method), exact controllability is equivalent to the following inequality:

$$\|(\varphi(0), \varphi_t(0))\|_{L^2(\Omega) \times H^{-1}(\Omega)}^2 \leq C \int_0^T \int_\omega \varphi^2 \, dx \, dt \quad (3.4)$$

for all solutions of the adjoint system (3.3). Note that this equivalence property is the analogue of the one (2.5) we have stated and proved in the previous section in the context of finite-dimensional systems.

This estimate is often also referred to as *continuous observability* since it provides a quantitative estimate of the norm of the initial data in terms of the observed quantity, by means of the observability constant C .

As we mentioned above, in contrast with the situation in finite-dimensional systems, for the observability inequality (3.4) to be true, it is not sufficient that the unique continuation property below holds:

$$\text{If } \varphi \equiv 0 \text{ in } \omega \times (0, T) \text{ then } \varphi \equiv 0. \quad (3.5)$$

Indeed, as we shall see, it may happen that the unique continuation property (3.5) holds, but the corresponding observed norm $[\int_0^T \int_\omega \varphi^2 \, dx \, dt]^{1/2}$ to be strictly weaker than the energy in $L^2(\Omega) \times H^{-1}(\Omega)$.

Inequality (3.4), when it holds, allows to estimate the total energy of the solution of (3.3) at time $t = 0$ by means of a measurement in the control region $\omega \times (0, T)$. But, in fact, the $L^2(\Omega) \times H^{-1}(\Omega)$ -energy is conserved in time, i.e.,

$$\|(\varphi(t), \varphi_t(t))\|_{L^2(\Omega) \times H^{-1}(\Omega)}^2 = \|(\varphi^0, \varphi^1)\|_{L^2(\Omega) \times H^{-1}(\Omega)}^2 \quad \forall t \in [0, T].$$

Thus, (3.4) is equivalent to

$$\|(\varphi^0, \varphi^1)\|_{L^2(\Omega) \times H^{-1}(\Omega)}^2 \leq C \int_0^T \int_{\omega} \varphi^2 \, dx \, dt, \quad (3.6)$$

or to

$$\int_0^T \|(\varphi(t), \varphi_t(t))\|_{L^2(\Omega) \times H^{-1}(\Omega)}^2 \, dt \leq C \int_0^T \int_{\omega} \varphi^2 \, dx \, dt. \quad (3.7)$$

When the observability inequality (3.4) holds the functional

$$\begin{aligned} J(\varphi^0, \varphi^1) &= \frac{1}{2} \int_0^T \int_{\omega} \varphi^2 \, dx \, dt \\ &\quad + \langle (\varphi(0), \varphi_t(0)), (u^1, -u^0) \rangle - \langle (\varphi^0, \varphi^1), (v^1, -v^0) \rangle \end{aligned} \quad (3.8)$$

has a unique minimizer $(\hat{\varphi}^0, \hat{\varphi}^1)$ in $L^2(\Omega) \times H^{-1}(\Omega)$ for all $(u^0, u^1), (v^0, v^1) \in H_0^1(\Omega) \times L^2(\Omega)$. The control $f = \hat{\varphi}$ with $\hat{\varphi}$ solution of (3.3) corresponding to the minimizer $(\hat{\varphi}^0, \hat{\varphi}^1)$ is such that the solution of (3.1) satisfies

$$u(T) = v^0, \quad u_t(T) = v^1. \quad (3.9)$$

The proof of this result is similar to the one of the finite-dimensional case we developed in the previous section. Thus we omit it. On the other hand, as in the finite-dimensional context, the controls we have built by minimizing J are those of minimal $L^2(\omega \times (0, T))$ -norm within the class of admissible controls.

But this observability inequality is far from being obvious and requires suitable geometric conditions on the control set ω , the time T and important technical developments.

Let us now discuss what is known about the observability inequality (3.6).

(a) *The method of multipliers.* Using multiplier techniques in the spirit of Morawetz [136], Ho in [92] proved that if one considers subsets of Γ of the form

$$\Gamma(x^0) = \{x \in \Gamma: (x - x^0) \cdot n(x) > 0\} \quad (3.10)$$

for some $x^0 \in \mathbb{R}^n$ (by $n(x)$ we denote the outward unit normal to Ω in $x \in \Gamma$ and by “ $\cdot$ ” the scalar product in $\mathbb{R}^n$) and if $T > 0$ is large enough, the following boundary observability inequality holds:

$$\|(\varphi(0), \varphi_t(0))\|_{H_0^1(\Omega) \times L^2(\Omega)}^2 \leq C \int_0^T \int_{\Gamma(x^0)} \left| \frac{\partial \varphi}{\partial n} \right|^2 d\Gamma dt \quad (3.11)$$

for all $(\varphi^0, \varphi^1) \in H_0^1(\Omega) \times L^2(\Omega)$.

This is the observability inequality that is required to solve the boundary controllability problem mentioned in Remark 3.1(g).

Later on inequality (3.11) was proved in [113] for any $T > T(x^0) = 2\|x - x^0\|_{L^\infty(\Omega)}$. This is the optimal observability time that one may derive by means of multipliers. We refer to [101] for a simpler derivation of the minimal time $T(x^0)$ with an explicit observability constant.

Let us recall that the method of multipliers relies on using $(x - x^0) \cdot \nabla \varphi$, φ and φ_t as multipliers in the adjoint system. Integrating by parts and combining the identities obtained in this way one gets (3.11).

Proceeding as in [113], vol. 1, Chapter VII, Section 2.3, one can easily prove that (3.11) implies (3.4) when ω is a neighborhood of $\Gamma(x^0)$ in Ω , i.e., $\omega = \Omega \cap \Theta$ where Θ is a neighborhood of $\Gamma(x^0)$ in $\mathbb{R}^n$, with $T > 2\|x - x^0\|_{L^\infty(\Omega \setminus \omega)}$. To do that it is sufficient to observe that the energy concentrated on the boundary that, in the context of Dirichlet boundary conditions, is fully determined by the L^2 -norm of the normal derivative $\partial \varphi / \partial n$, can be bounded above in terms of the energy on a neighborhood of that subset of the boundary. Thus, if the boundary observability inequality (3.11) is true, it should also hold when measurements are made on a neighborhood of the boundary. This provides an H^1 -version of (3.4). Inequality (3.4) itself can be obtained by a lifting argument based on taking time integrals of solutions.

Later on Osses in [139] introduced a new multiplier, which is basically a rotation of the previous one, obtaining a larger class of subsets of the boundary for which the observability inequality (3.11) holds.

It is important to underline that the situation in which the boundary observability inequality is obtained by the method of multipliers is limited by at least two reasons.

- The time T needs to be large enough. This is in agreement with the property of the finite speed of propagation underlying the wave model under consideration. But the method of multipliers rarely provides the optimal and minimal control time, some very particular geometries being excepted (for instance, the case of the ball, in which x^0 is taken to be its center).
- The geometry of the sets ω for which the inequality is proved using multipliers is very restrictive. One mainly recovers neighborhoods of subsets of the boundary of the form $\Gamma(x^0)$ as in (3.10) and these have a very special structure. In particular, when Ω is the square or a rectangle, the sets $\Gamma(x^0)$ are necessarily either two or three adjacent sides, or the whole boundary, depending on the location of x^0 . When Ω is a circle, $\Gamma(x^0)$ is always larger than a half-circumference. We refer to [130] where this issue

is addressed in more detail and some other limitations of the multiplier method both in what concerns control time and support of controls are proved. On the other hand, the multiplier method in itself does not give a qualitative justification for the need of such strict geometric restrictions. The microlocal approach we shall describe below provides a good insight into this issue and also shows that, in fact, controllability holds for a much larger class of subdomains ω .

There is an extensive literature on the use of multiplier techniques for the control and stabilization of wave-like equations. Here we have chosen to quote only some of the basic ones.

(b) *Microlocal analysis.* Bardos, Lebeau and Rauch [7] proved that, roughly, in the class of C^∞ domains, the observability inequality (3.4) holds if and only if (ω, T) satisfy the following *geometric control condition (GCC)* in Ω : *Every ray of Geometric Optics that propagates in Ω and is reflected on its boundary Γ enters ω in time less than T .*

To be more precise, [7] addresses the problem of boundary control which is technically more involved than the present one. The results in [7] apply however to the problem of interior control we address here in which the control acts in a subdomain ω .

In the formulation of the GCC above we have avoided some technical details related, in particular, with rays that get in contact with the boundary tangentially. Indeed, tangent rays may be diffractive or even enter the boundary. We refer to [7] for a deeper discussion of these issues, to [19] for a sharp necessary and sufficient condition and to [20] for the extension of this analysis to systems of PDE in which the notion of polarization of singularities plays an important role.

This complete characterization of the sets ω and times T for which observability holds provides also a good insight to the underlying reasons of the strict geometric conditions we encountered when applying multiplier methods. Roughly speaking, around each ray of Geometric Optics, on an arbitrarily small tubular neighborhood of it, one can concentrate solutions of the wave equation, the so-called Gaussian beams, that decay exponentially away from the ray. This suffices to show that, in case one of the rays does not enter the control region ω in a time smaller than T , the observability inequality may not hold. The construction of the Gaussian beams was developed by Ralston in [142] and [143]. In those articles the necessity of the GCC for observability was also pointed out. The main contribution in [7] was to prove that GCC is also sufficient for observability. As we mentioned above, the proof in [7] uses Microlocal Analysis and reduces the problem to show that the complete energy of solutions can be estimated uniformly provided all rays reach the observation subset in the given time interval. In fact, in [7] the more difficult problem of boundary controllability was addressed. This was done using the theory of propagation of singularities and a lifting lemma that allows getting estimates along the ray in a neighborhood of the boundary from the boundary estimate for the normal derivative.

This result was proved by means of Microlocal Analysis techniques. Recently the microlocal approach has been greatly simplified by Burq [16] by using the microlocal defect measures introduced by Gérard [78] in the context of the homogenization and the kinetic equations. In [16] the GCC was shown to be sufficient for exact controllability for domains Ω of class C^3 and equations with C^2 coefficients.

For the sake of completeness let us give the precise definition of *bicharacteristic ray*. Consider the wave equation with a scalar, positive and smooth variable coefficient $a = a(x)$,

$$\varphi_{tt} - \operatorname{div}(a(x)\nabla\varphi) = 0. \quad (3.12)$$

Bicharacteristic rays solve the Hamiltonian system

$$\begin{cases} x'(s) = -a(x)\xi, & t'(s) = \tau, \\ \xi'(s) = \nabla a(x)|\xi|^2, & \tau'(s) = 0. \end{cases} \quad (3.13)$$

Rays describe the microlocal propagation of energy. The projections of the bicharacteristic rays in the (x, t) variables are the rays of Geometric Optics that enter in the GCC.² As time evolves the rays move in the physical space according to the solutions of (3.13). Moreover, the direction in the Fourier space (ξ, τ) in which the energy of solutions is concentrated as they propagate is given precisely by the projection of the bicharacteristic ray in the (ξ, τ) variables. When the coefficient $a = a(x)$ is constant all rays are straight lines and carry the energy outward, which is always concentrated in the same direction in the Fourier space, as expected.

But for variable coefficients the dynamics is more complex and can lead to some unexpected phenomena [123]. GCC is still a sufficient and almost necessary condition for observability to hold. But one has to keep in mind that, in contrast with the situation of the constant coefficient wave equation, for variable coefficients, some rays may never reach the exterior boundary. There are for instance wave equations with smooth coefficients for which there are periodic rays that never meet the exterior boundary. Thus, the case in which ω is a neighborhood of the boundary of the domain Ω , for which observability holds for the constant coefficient wave equation, does not necessarily fulfill the GCC for variable coefficients. In those cases boundary observability fails. Our intuition is often strongly inspired on the constant coefficient wave equation for which all rays are straight lines tending to infinity at a constant velocity, which, in particular, implies that the rays will necessarily reach the exterior boundary of any bounded domain. But for variable coefficients rays are not straight lines any more and the situation may change drastically. We refer to [123] for a discussion of this issue. We also refer to the article by Miller [130] where this problem is analyzed from the point of view of “escape functions”, a sort of Lyapunov functional allowing to test whether all rays tend to infinity or not.

On the other hand, this Hamiltonian system (3.13) describes the dynamics of rays in the interior of the domain where the equation is satisfied. But when rays reach the boundary they are reflected according to the laws of Geometric Optics.

So far the microlocal approach is the one leading to the sharpest observability results, in what concerns the geometric requirements on the subset ω where the control is applied and on the control time, but it requires more regularity of coefficients and boundaries than multipliers do. The drawback of the multiplier method, which is much simpler to apply, is that it only works for restricted classes of wave equations and that it does not give sharp

²This is rigorously true in the interior of the domain. But, to take boundary effects into account, one has to define the so-called generalized bicharacteristic rays (see [7]).

results as those that the Geometric Optics interpretation of the observability inequality predicts.

The multiplier approach was adapted to the case of nonsmooth domains by Grisvard in [82] and the microlocal one by Burq in [16] and [17].

(c) *Carleman inequalities.* The third most common and powerful technique to derive observability inequalities is based on the so called Carleman inequalities. It can be viewed as a more developed version of the classical multiplier technique. It applies to a wide class of equations with variable coefficients, under less regularity conditions than the microlocal approach requires. The Carleman approach needs, roughly, that the coefficients of the principal part be Lipschitz continuous. Thus, with respect to the method of multipliers, the Carleman approach has the advantage of being more flexible and allowing to address variable coefficients, and, with respect to the microlocal one, that it requires less regularity on coefficients and domain.

But one of the major advantages of this approach is that it allows considering, for instance, lower-order perturbations and getting explicit bounds on the observability constant in terms of the potentials entering in it. We refer for instance to [95] and [164]. This is particularly important when dealing with nonlinear problems by means of linearization and fixed point arguments (see [168]). More generally, the Carleman inequality approach provides explicit bounds of the observability constant for systems depending on an extra parameter. For instance, in [121] the Carleman inequalities play a key role when deriving the continuity of controls for the following singular perturbation problem connecting dissipative wave equations with the heat equation

$$\varepsilon u_{tt} - \Delta u + u_t = f \mathbb{1}_\omega,$$

and in [37] where the same issue is addressed for the convective equation with vanishing viscosity and control in the interior or on the boundary,

$$u_t + u_x - \varepsilon u_{xx} = 0.$$

Note however that, as in the microlocal approach, the Carleman inequalities do not yield observability inequalities for all variable coefficients and that, in fact, as we said above, they fail to be true unless appropriate assumptions on coefficients are made. The way this is seen at the level of Carleman inequalities is not as explicit as in the microlocal approach where the methodology is based on the ray analysis. But for Carleman inequalities to yield observability estimates, suitable weight functions are needed, and this requires of some assumptions on the coefficients and its first-order derivatives, as it is the case for multiplier methods to apply [162]. These assumptions on the coefficients needed in the context of Carleman inequalities and multiplier techniques imply, in particular, that rays reach the observation region ω (see [130]). The converse is not true as proved in [130].

(d) *Spectral estimates.* More recently an interesting characterization of the observability inequality in terms of the spectrum of the underlying operator has been derived in [21, 118, 132] and [144]. The result reads essentially as follows. Let $\{\psi_k\}_{k \geq 1}$ be an orthonormal

basis of $L^2(\Omega)$ constituted by the eigenfunctions of the Dirichlet Laplacian and let $\{\lambda_k\}_{k \geq 1}$ be the corresponding eigenvalues

$$\begin{cases} -\Delta \psi_j = \lambda_j \psi_j & \text{in } \Omega, \\ \psi_j = 0 & \text{on } \partial\Omega. \end{cases} \quad (3.14)$$

Then the observability inequality (3.4) holds for the wave equation if and only if the eigenfunctions satisfy the following property: *there exist some $\sigma > 0$ and $C > 0$ such that*

$$\sum_{I_k} |a_j|^2 \leq C \int_{\omega} \left| \sum_{I_k} a_j \psi_j(x) \right|^2 dx \quad \forall k \geq 1, \forall a_j \in \mathbb{R}, \quad (3.15)$$

where the sums run over the sets of indexes I_k as follows:

$$I_k = \{j: |\sqrt{\lambda_j} - \sqrt{\lambda_k}| \leq \sigma\}. \quad (3.16)$$

By analogy with the 1- d case we shall discuss in the following subsection, one could expect that, under this spectral condition (3.15), the time T needed for the observability estimate (3.4) to hold to be $2\pi/\sigma$. Whether this is true or not is an open problem. We refer to [159], Theorem 6.4.5, Section 6.3, for an explicit estimate on the time needed for this to hold. But that estimate is significantly larger than $2\pi/\sigma$.

As we shall see in the following subsection, in 1- d , this optimal time can be achieved. But this requires the use of the classical Ingham inequality and of a spectral gap condition. The latter is not fulfilled for the wave equation in several space dimensions. When the gap condition holds, the set of indices I_k is reduced to $\{k\}$ and then the eigenfunction estimate (3.15) is reduced to check it for isolated eigenfunctions.

The characterization above reduces the problem of observability of the wave equation to the obtention of the estimate (3.15) for linear combinations of eigenfunctions.³ But, of course, the proof of the latter is in general a noneasy task. It requires, once more, of suitable geometric assumptions on the subdomain ω where the control is concentrated and can be developed by methods similar to those we described above for addressing directly the observability inequality for the wave equation. In particular, multipliers and Carleman inequalities may be used. But in the context of the condition (3.15) we can play with the advantage of using, for instance, multipliers that depend on the leading frequency of the wave packet under consideration. The two approaches, the dynamic one that consists in addressing directly the evolution equation (3.3) and the spectral one, end up giving similar results. However, because of its very nature, in the context of the wave equation, in order to obtain the control under the sharp GCC, in a way or another, the microlocal analysis needs to be used, even if it is in the obtention of (3.15).

In any case this spectral characterization of the observability inequality is of interest since, to some extent, it provides a natural extension of the methods based on nonharmonic Fourier series techniques and Ingham inequalities that so successfully apply in 1- d problems and that we describe now.

³Note however that, as indicated as above, this characterization does not seem to provide the optimal time.

This method provides also a new easy proof of the fact that the controllability of the Schrödinger equation in an arbitrarily small time can be derived as consequence of the controllability of the wave equation in some finite time.

3.3. Ingham inequalities and Fourier series techniques

We have described here the HUM, which reduces the control problem to an observability one for the adjoint wave equation, and some tools to prove observability inequalities. There are other techniques that are also useful to address the problem of observability. In particular the theory of nonharmonic Fourier series and the so-called Ingham inequality allows obtaining sharp observability inequalities for a large class of one-dimensional wave-like equations. As we mentioned above, the Ingham inequality [99,163] plays, in 1- d , a similar role to the spectral characterization (3.15) above. It reads as follows.

THEOREM 3.1 (Ingham's theorem [99]). *Let $\{\mu_k\}_{k \in \mathbb{Z}}$ be a sequence of real numbers such that*

$$\mu_{k+1} - \mu_k \geq \gamma > 0 \quad \text{for all } k \in \mathbb{Z}. \quad (3.17)$$

Then, for any $T > 2\pi/\gamma$ there exists a positive constant $C(T, \gamma) > 0$ such that

$$\frac{1}{C(T, \gamma)} \sum_{k \in \mathbb{Z}} |a_k|^2 \leq \int_0^T \left| \sum_{k \in \mathbb{Z}} a_k e^{i\mu_k t} \right|^2 dt \leq C(T, \gamma) \sum_{k \in \mathbb{Z}} |a_k|^2 \quad (3.18)$$

for all sequences of complex numbers $\{a_k\} \in \ell^2$.

REMARK 3.2.

1. Although the most common use of Ingham's theorem is precisely the inequality (3.18), in the original article by Ingham [100], it was also proved that, under the same gap condition, there exists a constant $C(T, \gamma) > 0$ such that the following L^1 -version is also true:

$$|a_n| \leq C(T, \gamma) \int_0^T \left| \sum_{k \in \mathbb{Z}} a_k e^{i\mu_k t} \right| dt \quad (3.19)$$

for all $n \in \mathbb{Z}$. In fact, as proved by Ingham, the constant $C(T, \gamma)$ can be taken to be the same in (3.19) and in its L^2 -analogue in (3.18).

2. The original Ingham inequality was proved under the gap condition (3.17). However it is by now well known that this gap condition can be weakened, extending the range of possible applications. It is for instance well known that for the inequalities (3.18) to be true it suffices that all eigenvalues are distinct and that the gap condition is fulfilled asymptotically for high frequencies. Under this asymptotic gap condition the time T for the first inequality in (3.18) is the same, γ being the asymptotic gap. We refer for instance

to [125] where an explicit estimate on the constants in (3.18) is given. These constants depend on the global gap, on the asymptotic gap and also on how rapidly this asymptotic gap is achieved. We underline however that the time T only depends on the asymptotic gap. Further generalizations have also been given. We refer for instance to the book [102] which contains an extension of that inequality covering families of eigenfrequencies which are, roughly, a finite union of sequences, satisfying each of them separately a gap condition. This kind of generalization is very useful in particular when dealing with networks of vibrations (see [41]).

Let us develop the details of the application of the Ingham inequality for the 1- d wave equation to better explain the connection with observability.

Consider the 1- d domain $\Omega = (0, \pi)$ and the adjoint wave equation with Dirichlet boundary conditions

$$\begin{cases} \varphi_{tt} - \varphi_{xx} = 0 & \text{in } (0, \pi) \times (0, T), \\ \varphi(x, t) = 0 & \text{for } x = 0, \pi, t \in (0, T), \\ \varphi(x, T) = \varphi^0(x), \quad \varphi_t(x, T) = \varphi^1(x) & \text{in } (0, \pi). \end{cases} \quad (3.20)$$

Consider any nonempty subinterval ω of Ω as observation region.

The solutions of this wave equation can be written in Fourier series in the form

$$\varphi(x, t) = \sum_{k \in \mathbb{Z}} a_k e^{ikt} \sin(kx).$$

When applying Ingham's inequality for this series the relevant eigenfrequencies are $\mu_k = k$. The gap condition (3.17) is then clearly satisfied in this case with $\gamma = 1$. Thanks to (3.18), obtaining observability estimates for the solutions of the wave equation, when $T > 2\pi$, can be reduced to the obtention of similar estimates for the eigenfunctions. More precisely, if $T > 2\pi$, (3.4) holds for the solutions of the 1- d wave equation because the eigenfunctions $\{\sin(kx)\}_{k \geq 1}$ satisfy

$$\int_{\omega} \sin^2(kx) dx \geq c_{\omega} \quad \forall k \geq 1. \quad (3.21)$$

The last condition is easy to obtain since $\sin^2(kx)$ converges weakly to $1/2$ in $L^2(0, \pi)$ as $k \rightarrow \infty$.

Indeed, applying Fubini's lemma and Ingham's inequality for all $x \in \omega$ to the series $\varphi(x, t) = \sum_{k \in \mathbb{Z}} a_k e^{ikt} \sin(kx)$, viewing $a_k \sin(kx)$ as coefficients and applying (3.21) we deduce that

$$\begin{aligned} \int_0^T \int_{\omega} \varphi^2 dx dt &= \int_0^T \int_{\omega} \left| \sum_{k \in \mathbb{Z}} a_k e^{ikt} \sin(kx) \right|^2 dx dt \\ &= \int_{\omega} \int_0^T \left| \sum_{k \in \mathbb{Z}} a_k e^{ikt} \sin(kx) \right|^2 dt dx \end{aligned}$$

$$\geq C \int_{\omega} \sum_{k \in \mathbb{Z}} |a_k|^2 \sin^2(kx) \, dx \geq C \sum_{k \in \mathbb{Z}} |a_k|^2. \quad (3.22)$$

This proves (3.4) because $\sum_{k \in \mathbb{Z}} |a_k|^2$ is equivalent to the $(L^2 \times H^{-1})$ -norm of the initial data of (3.3). In other words, (3.4) holds.

Note that in the application to the 1- d wave equation on $(0, 1)$, the Ingham inequality does not yield the observability inequality for $T = 2$ but rather only for $T > 2$. In fact it is well known that the Ingham inequality does not hold in general for the optimal time $T = 2\pi/\gamma$ but only for $T > 2\pi/\gamma$ (see [163]). But, in the present case, because of the orthogonality of trigonometric polynomials in $L^2(0, 2)$ the estimate holds for $T = 2$ too. In fact, in the case (3.20), if the control subdomain is the subinterval $\omega = (a, b)$ of $(0, \pi)$, the minimal control time is $2 \max(a, \pi - b)$.

Let us now comment on the relation between the Ingham inequality approach described here for 1- d problems and the spectral characterization (3.15) developed in the previous section. Note that, because of the gap condition, whenever $\sigma < \gamma = 1$, the sets I_k in (3.15) are reduced to the single eigenvalue λ_k . Consequently, the inequality in (3.15) reduces to (3.21) that, as we have seen, trivially holds.

Consequently the spectral characterization (3.15) plays a similar role as the Ingham inequality, but in any space dimension. However, as we mentioned above, the use of the spectral condition (3.15), (3.16) in several space dimensions is much more subtle since it forces us to deal with wave packets, while in 1- d , one has only to check the uniform observability of individual eigenfunctions as in (3.21). Note that Ingham's inequality cannot be applied for the wave equation in multidimensional problems since they grow asymptotically as $\lambda_j \sim c(\Omega)j^{2/n}$ and the gap vanishes. Despite this fact the observability inequality (3.4) may hold under suitable geometric conditions on ω and for sufficiently large values of time T . This is precisely, as the spectral condition (3.15) indicates, because of the uniform observability of the spectral wave packets.

The Ingham inequality approach can be applied to a variety of 1- d problems in which the Fourier representation of solutions can be used (mainly when the coefficients are time-independent) provided the gap condition holds. In this way one can address wave equations with variable coefficients, Airy equations, beam and Schrödinger equations, etc. Ingham inequality is also useful to address problems in which the control is localized on an isolated point or other singular ways, situations that cannot be handled by multipliers, for instance, see [88] and [158]. We refer to [127] for a brief introduction to this subject and to the monographs by Avdonin and Ivanov [6] and Komornik [101] and that by Komornik and Loreti [102] for a more complete presentation and discussion of this approach, intimately related also to the moment problem formulation of the control problem. We refer to [150] for a discussion of the moment problem approach. We also refer to the book [41] for an application of nonharmonic Fourier series methods to the control of waves on networks.

It is also important to underline that both, the Ingham approach and the spectral characterization (3.15) apply only for equations allowing a spectral decomposition of solutions. Thus, for instance, it does not apply to wave equations with coefficients depending both on x and t , or with lower-order potentials of the form

$$\varphi_{tt} - \Delta \varphi + a(x, t)\varphi = 0.$$

This is certainly the main drawback of the Fourier approach to observability of solutions of wave equations.

3.4. Approximate controllability

So far we have analyzed the problem of exact controllability. Let us now briefly discuss the *approximate controllability problem*.

According to the definition above (see Section 3.1), the problem of approximate controllability is equivalent to being able to find controls f for all pairs of initial and final data in the energy space $H_0^1(\Omega) \times L^2(\Omega)$, such that the following holds:

$$\|(u(T) - v^0, u_t(T) - v^1)\|_{H_0^1(\Omega) \times L^2(\Omega)} \leq \varepsilon. \quad (3.23)$$

Note that the property of approximate controllability guarantees that one can drive the state of the system arbitrarily close to the final target (v^0, v^1) . However, it does not ensure in itself that we can reach the final target exactly, i.e., that we can take $\varepsilon = 0$ in (3.23).

In fact, the value $\varepsilon = 0$ is reached when exact controllability holds, in which case there exists a control f such that the solution satisfies exactly (3.9). In this case the approximate controls f_ε are uniformly bounded (with respect to ε) and, as ε tends to zero, converge to an exact control.

In finite dimension, approximate controllability and exact controllability are equivalent notions. But this is no longer the case in the context of PDE because of the intrinsic infinite-dimensional nature of the state space. Indeed, in infinite-dimensional spaces there are strict dense subspaces, while in finite dimension they do not exist.

This is particularly important in the context of the wave equation. As we have seen, for the exact controllability property to hold, one needs to impose rather strict geometric conditions on the control set. However, as we shall see, these restrictions are not needed for approximate controllability. On the other hand, approximate controllability is relevant from the point of view of applications in which the notion of exact controllability might seem to introduce a too strong constraint on the final state. However, as we mentioned above, when exact controllability fails, the size of this control diverges typically exponentially as ε tends to zero (see [106,147]). This is an important warning about the effective use of the property of approximate controllability when exact controllability fails.

The approximate controllability property is equivalent to a *unique continuation* one for the adjoint system

$$\begin{cases} \varphi_{tt} - \Delta \varphi = 0 & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x), \quad \varphi_t(x, T) = \varphi^1(x) & \text{in } \Omega. \end{cases} \quad (3.24)$$

More precisely, system (3.1) is approximately controllable if and only if the following holds:

$$\varphi \equiv 0 \quad \text{in } \omega \times (0, T) \implies (\varphi^0, \varphi^1) \equiv (0, 0). \quad (3.25)$$

This unique continuation property is the analogue of that in (2.13) arising in the finite-dimensional theory.

By using Holmgren's uniqueness theorem (see [93]) it can be easily seen that (3.25) holds if T is large enough, under the sole condition that ω is a nonempty open subset of Ω . We refer to [113], Chapter 1, for a discussion of this problem. At this respect it is important to underline that, even if the unique continuation property holds for all subdomains ω , for the observability inequality to be true ω is required to satisfy the GCC.

Let us now, assuming that the uniqueness property (3.25) holds (we shall return to this issue later on), analyze how approximate controllability can be obtained out of it. There are at least two ways of checking that (3.25) implies the approximate controllability property.

- (a) The application of Hahn–Banach theorem.
- (b) The variational approach developed in [115].

We refer to [127] for a presentation of these methods.

In fact, when approximate controllability holds, then the following (apparently stronger) statement also holds.

THEOREM 3.2 [175]. *Let E be a finite-dimensional subspace of $H_0^1(\Omega) \times L^2(\Omega)$ and let us denote by π_E the corresponding orthogonal projection. Then, if approximate controllability holds, or, equivalently, if the unique continuation property (3.25) is satisfied, for any $(u^0, u^1), (v^0, v^1) \in H_0^1(\Omega) \times L^2(\Omega)$ and $\varepsilon > 0$ there exists $f_\varepsilon \in L^2(Q)$ such that the solution u_ε of (3.1) satisfies*

$$\begin{aligned} & \| (u_\varepsilon(T) - v^0, u_{\varepsilon,t}(T) - v^1) \|_{H_0^1(\Omega) \times L^2(\Omega)} \leq \varepsilon, \\ & \pi_E(u_\varepsilon(T), u_{\varepsilon,t}(T)) = \pi_E(v^0, v^1). \end{aligned} \quad (3.26)$$

This result, that will be referred to as the *finite-approximate controllability property*, may be proved in several ways. But, in particular, it can be obtained easily by a suitable modification of the variational approach introduced in [115] that we shall describe at the end of this subsection. This variational approach, in all cases, provides the control of minimal L^2 -norm within the class of admissible controls. This makes the method particularly interesting and robust.⁴

The functional to be minimized to get approximate controllability is as follows:

$$\begin{aligned} J_\varepsilon(\varphi^0, \varphi^1) &= \frac{1}{2} \int_0^T \int_\omega \varphi^2 \, dx \, dt + \varepsilon \|(\varphi^0, \varphi^1)\|_{L^2(\Omega) \times H^{-1}(\Omega)} \\ &\quad + \langle (\varphi(0), \varphi_t(0)), (u^1, -u^0) \rangle - \langle (\varphi^0, \varphi^1), (v^1, -v^0) \rangle. \end{aligned} \quad (3.27)$$

⁴Robustness is one of the key requirements in control theoretical applications. Indeed, although the control strategy is built on the basis of some specific modeling (or *plant* in the engineering terminology), in practice, due to uncertainty or to the intrinsic inaccuracies of the model considered, one needs to be sure that the control will also work properly under those unavoidable perturbations such as measurement noise and external disturbances. Robustness is then a fundamental requirement to be fulfilled by the control mechanism. The advantage of building and using controls that come out of a variational principle, by minimizing a suitable quadratic, convex and coercive functional, is that, slight changes of the functional will produce a smooth behavior on the control, as can be proved by the classical techniques in Γ -convergence theory [42]. This, of course, has to be carefully checked in each particular case, but it is a methodology that works satisfactorily well in most cases.

When adding the term $\varepsilon|(\varphi^0, \varphi^1)|_{L^2(\Omega) \times H^{-1}(\Omega)}$ in the functional to be minimized, the corresponding Euler–Lagrange equations for the minimizers turn out to be (3.23) instead of (3.9), which corresponds to the property of exact controllability and to the minimization of the functional J in (3.8). Consequently, this added term acts as a regularization of the functional J and, consequently, relaxes the controllability property obtained as a direct consequence of the optimality condition satisfied by the minimizers. It is also interesting to observe that, when adding this term to the functional, its coercivity is much easier to derive since it holds as a direct consequence of the unique continuation property (3.5) of the adjoint system, without requiring the observability inequality (3.4) to hold. Note however that this argument does not give any information on the size of the control needed to reach the target up an ε -distance.

Similarly, the finite-approximately controllability property can be achieved as a consequence of the unique continuation property by minimizing the functional

$$J_{\varepsilon,E}(\varphi^0, \varphi^1) = \frac{1}{2} \int_0^T \int_{\omega} \varphi^2 \, dx \, dt + \varepsilon \|(I - \pi_E^*)(\varphi^0, \varphi^1)\|_{L^2(\Omega) \times H^{-1}(\Omega)} + \langle (\varphi(0), \varphi_t(0)), (u^1, -u^0) \rangle - \langle (\varphi^0, \varphi^1), (v^1, -v^0) \rangle. \quad (3.28)$$

Note that the main difference between the functionals J_{ε} and $J_{\varepsilon,E}$ is that in the latter the relaxation term is weaker since we only add the norm of the projection $I - \pi_E^*$ of the data of the adjoint system and not the full norm. It is important to observe that, in the proof of the coercivity of the functional $J_{\varepsilon,E}$, the fact that the operator π_E^* is compact plays a key role. For this reason the space E is assumed to be of finite dimension. The projection π_E^* is defined by duality as

$$\langle \pi_E^*(\varphi^0, \varphi^1), (u^1, -u^0) \rangle = \langle (\varphi^0, \varphi^1), \pi_E(u^1, -u^0) \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $L^2(\Omega) \times H^{-1}(\Omega)$ and its dual $L^2(\Omega) \times H_0^1(\Omega)$.

The functional $J_{\varepsilon,E}$ can be obtained by a duality argument in order to get the control of minimal norm among the admissible ones. We refer for instance to [112] that addresses this issue in the context of the exact controllability of the wave equation. To be more precise, the control of minimal norm f is characterized by the minimality condition (see [112]),

$$f \in \mathcal{U}_{\text{ad}}: \quad \|f\|_{L^2(\omega \times (0,T))} = \min_{g \in \mathcal{U}_{\text{ad}}} \|g\|_{L^2(\omega \times (0,T))}, \quad (3.29)$$

where $\mathcal{U}_{\text{ad}}$ is the set of admissible controls. More precisely,

$$\mathcal{U}_{\text{ad}} = \{f \in L^2(\omega \times (0, T)): \text{ the solution } u \text{ of (3.1) satisfies (3.26)}\}. \quad (3.30)$$

The dual, in the sense of Fenchel–Rockafellar, of this minimization problem turns out to be precisely that of the minimization of the functional $J_{\varepsilon,E}$ with respect to (φ^0, φ^1) in $L^2(\Omega) \times H^{-1}(\Omega)$.

In all these cases the control f we are looking for is the restriction to ω of the solution of the adjoint system (3.24) with the initial data that minimize the corresponding functional.

The results above on uniqueness for the adjoint wave equation and approximate controllability of the state equation hold for wave equations with analytic coefficients too. Indeed, the approximate control problem can be reduced to the unique continuation one and the latter may be solved by means of Holmgren's uniqueness theorem when the coefficients of the equation are analytic. However, the problem is still not completely solved in the frame of the wave equation with lower-order potentials $a \in L^\infty(Q)$ of the form

$$u_{tt} - \Delta u + a(x, t)u = f \mathbb{1}_\omega \quad \text{in } Q. \quad (3.31)$$

Once again the problem of approximate controllability of this system is equivalent to the unique continuation property of its adjoint. We refer to Alinhac [2], Tataru [157] and Robbiano and Zuily [148] for deep results in this direction. But, roughly, we can say that in the class of bounded coefficients $a = a(x, t)$ we still do not have local sharp results on unique continuation allowing to handle equations of the form (3.31) in full generality. The existing ones require either some analyticity properties of the coefficients [148, 157] or some geometric constraints in ω to apply the Carleman inequalities techniques [164]. On the other hand, it is well known that the unique continuation property may fail in general [2]. A complete picture is still to be found.

3.5. Quasibang-bang controls

In the finite-dimensional setting we have shown that, by slightly changing the functional to be minimized to get the controls, one can build bang-bang controls.

There is a very natural way of adapting this idea in the context of the wave equation. Indeed, essentially, it consists in replacing the functional (3.27) by its L^1 -version

$$J_{\text{bb}, \varepsilon}(\varphi^0, \varphi^1) = \frac{1}{2} \left[\int_0^T \int_\omega |\varphi| \, dx \, dt \right]^2 + \varepsilon \|(\varphi^0, \varphi^1)\|_{L^2(\Omega) \times H^{-1}(\Omega)}^2 \\ + \langle (\varphi(0), \varphi_t(0)), (u^1, -u^0) \rangle - \langle (\varphi^0, \varphi^1), (v^1, -v^0) \rangle. \quad (3.32)$$

This functional, as we shall see, is motivated by the search of controls of minimal L^∞ -norm too.

It is convex and continuous in the space $L^2(\Omega) \times H^{-1}(\Omega)$. It is also coercive as a consequence of the unique continuation property (3.5). Therefore, a minimizer exists. Let us denote it by $(\hat{\varphi}^0, \hat{\varphi}^1)$. One can then see that there exists a quasi bang-bang control $f \in \int_0^T \int_\omega |\hat{\varphi}| \, dx \, dt \operatorname{sgn}(\hat{\varphi})$, $\hat{\varphi}$ being the solution of the adjoint system corresponding to the minimizer, such that the approximate controllability condition (3.23) holds. Note however that the bang-bang structure of the control is not guaranteed. Indeed, the fact that $f \in \int_0^T \int_\omega |\hat{\varphi}| \, dx \, dt \operatorname{sgn}(\hat{\varphi})$ means that $f = \pm \int_0^T \int_\omega |\hat{\varphi}| \, dx \, dt$ in the set in which $\hat{\varphi} \neq 0$, but simply that $f \in [-\int_0^T \int_\omega |\hat{\varphi}| \, dx \, dt, \int_0^T \int_\omega |\hat{\varphi}| \, dx \, dt]$, in the set where $\hat{\varphi} = 0$. Obviously, one

cannot exclude the null set $\hat{\varphi} \equiv 0$ to be large. As we have seen, in the context of finite-dimensional systems, because of the analyticity of solutions, this null set is reduced to be a finite number of switching times in which the bang-bang control changes sign. But for the wave equation one cannot exclude it to be even a nonempty and open subset of $\omega \times (0, T)$. In fact the explicit computations in [85] show that, for the one-dimensional wave equation, the set of reachable states by means of bang-bang controls is rather restricted. This is natural to be expected and it is particularly easy to understand in the case of boundary control. Indeed, according to D'Alembert's formula, the effect of the boundary control in the state, solution of the 1- d wave equation, is roughly that of reproducing at time $t = T$ the structure of the controller (assuming we start from the null initial datum). Thus, if the control is of bang-bang form and, in particular, piecewise constant, that necessarily imposes a very simple geometry of the reachable functions.

A complete analysis of whether the quasibang-bang controls we have obtained above are bang-bang or not and a characterization of the set of initial data for which the bang-bang controls exist in the multidimensional case is still to be done.

It is also worth noting that the same problem of the existence of quasibang-bang controls was investigated in [84] in the context of approximate controllability of the 1- d wave equation but by replacing the energy space $H_0^1(\Omega) \times L^2(\Omega)$ by its L^∞ -version $W_0^{1,\infty}(\Omega) \times L^\infty(\Omega)$. In this case it was proved that relaxation occurs and that the controls that are obtained are not longer of quasibang-bang form. This is due to the fact that, when addressing this problem by the variational tools we have developed, one needs to modify the functional $J_{bb,\varepsilon}$ above by replacing the added term $\varepsilon|(\varphi^0, \varphi^1)|_{L^2(\Omega) \times H^{-1}(\Omega)}$ by its L^1 -version. This makes the problem of minimization not to be formulated in a reflexive Banach space. Relaxation phenomena may not be excluded a priori and, in fact, as the explicit examples in [84] show, they occur making the minimizers to develop singularities and, eventually, making the controls obtained in this way not to be of quasibang-bang form.

The same problem can be considered in the case of exact controllability in which the functional $J_{bb,\varepsilon}$ has to be replaced by J_{bb} ,

$$J_{bb}(\varphi^0, \varphi^1) = \frac{1}{2} \left[\int_0^T \int_\omega |\varphi| \, dx \, dt \right]^2 + \langle (\varphi(0), \varphi_t(0)), (u^1, -u^0) \rangle - \langle (\varphi^0, \varphi^1), (v^1, -v^0) \rangle. \quad (3.33)$$

The functional setting in which this functional has to be minimized is much less clear. In principle, one should work in the class of solutions of the adjoint system for which $\varphi \in L^1(\omega \times (0, T))$. Then, following the HUM, it would be natural to consider the functional J_{bb} as being defined in the Banach space defined as completion of test functions $(\varphi^0, \varphi^1) \in \mathcal{D}(\Omega) \times \mathcal{D}(\Omega)$ with respect to the norm

$$\int_0^T \int_\omega |\varphi| \, dx \, dt.$$

The characterization of the space X is a difficult open problem. Under the assumption that ω satisfies the GCC, if T is large enough, using the known observability estimates in

energy spaces and Sobolev embeddings, one could show that X is continuously embedded in an energy space of the form $H^s(\Omega) \times H^{s-1}(\Omega)$ for some negative $s < 0$. But a complete characterization of the space X is certainly extremely hard to get, except, maybe, in 1- d , in which the D'Alembert formula holds. But even in 1- d this remains to be done. Note also that, in 1- d the L^1 -version of the Ingham inequality (3.18) could also help. In particular, that would imply that the space X is continuously embedded in the space of solutions of the wave equation with Fourier coefficients in ℓ^∞ . But, from this point of view, a complete characterization of X is also unknown.

But, regardless of what the exact characterization of X is, the space X turns out to be nonreflexive. Thus, the minimization problem is not guaranteed to have a solution. Relaxation phenomena could occur, and one could be obliged to work in the space of solutions of the wave equation that are bounded measures when restricted to $\omega \times (0, T)$. In [84] this relaxation process has been shown to arise in the context of the boundary control of the wave equation. The explicit results in [84] for the 1- d wave equation show that, in general, the controls are not of bang-bang form. The existence of the minimizer of J_{bb} in X and the regularity of the minimizers constitute then interesting open problems.

A complete analysis of this minimization problem and its connections with L^∞ -minimal norm controls and its bang-bang structure is to be developed for multidimensional problems.

The same questions arise in the context of boundary control.

REMARK 3.3. Obviously, there is a one-parameter family of L^p variational problems making the link between the L^2 -optimal controls considered in the previous section and the quasibang-bang controls that we have analyzed above. Indeed, for instance, in the context of exact controllability (the same could be said about approximate controllability and finite-approximate controllability), one can consider the L^p -version of the functional to be minimized

$$J_p(\varphi^0, \varphi^1) = \frac{1}{2} \left[\int_0^T \int_\omega |\varphi|^p dx dt \right]^{2/p} + \langle (\varphi(0), \varphi_t(0)), (u^1, -u^0) \rangle - \langle (\varphi^0, \varphi^1), (v^1, -v^0) \rangle. \quad (3.34)$$

This functional has to be analyzed in the corresponding Banach space X_p of solutions of the adjoint heat equation whose restrictions to $\omega \times (0, T)$ belongs to L^p . In this case the space is reflexive and the minimizers exist. The corresponding controls take the form

$$f = \left[\int_0^T \int_\omega |\hat{\varphi}|^p dx dt \right]^{2/p-1} |\hat{\varphi}|^{p-2} \hat{\varphi},$$

where $\hat{\varphi}$ is the solution of the adjoint system corresponding to the minimizer. When $p = 2$ this corresponds to the L^2 -control, which is known to exist under the GCC. When $p = 1$ it corresponds to the quasibang-bang controls we have discussed. This justifies the use of the notation J_{bb} for J_1 . Analyzing the behavior of the space X_p and of the minimizers as p goes from 2 to 1 is an interesting open problem.

The functional J_p can also be obtained by the Fenchel–Rockafellar duality principle when searching controls of minimal $L^{p'}$ -norm.

3.6. Stabilization

In this article we have not addressed the problem of stabilization. Roughly, in the context of linear equations, it can be formulated as the problem of producing the exponential decay of solutions through the use of suitable dissipative feedback mechanisms.

The main differences with respect to controllability problems are that:

- in controllability problems the time varies in a finite interval $0 < t < T < \infty$, while in stabilization problems, the time t tends to infinity;
- in control problems, the control can enter in the system freely in an open-loop manner, while in stabilization the control is of feedback or closed-loop form.

As in the context of controllability there are several degrees of stability or stabilizability of a system that are of interest.

- One can simply analyze the decay of solutions. This is typically done using LaSalle's invariance principle (see [87]). At the level of controllability this would correspond to a situation in which approximate controllability holds. Indeed, both problems are normally reduced to proving an unique continuation property. Still, the one corresponding to stabilization is normally easier to deal with since the time t varies in the whole $(0, \infty)$ and, because of this, often, the unique continuation problem is reduced to analyzing it at the spectral level.
- The more robust and strong stability property one can look for is that in which the energy of solutions (the norm of the solution in the state-space) tends to zero exponentially uniformly as $t \rightarrow \infty$. This normally requires of very efficient feedback mechanisms and, at the level of controllability, corresponds to the property of exact controllability. At this respect it is important to note that for linear dissipative semigroups, if the norm of the semigroup tends to zero, it necessarily decays exponentially. Thus the only uniform decay property that makes sense for linear semigroups is the exponential one. Obviously, that is not the case for equations with nonlinear damping terms in which the decay rate may be polynomial, logarithmic, or of any other order, depending of the strength of the nonlinearity appearing in the feedback law (see, for instance, [89,119,171]).
- An interesting intermediate situation is that in which the uniform exponential decay fails but one is able to prove the polynomial decay of the solutions in the domain of the generator of the underlying semigroup. This normally corresponds, at the control level, to situations in which the control mechanism is unable to yield exact controllability properties, but guarantees the controllability of all data with slightly stronger regularity properties. That is for instance the case in the context of multistructures as those considered in [86] where a model for the vibrations of strings coupled with point masses is considered and in [41] where wave equations on graphs are addressed.

But not only these two properties (controllability/stabilizability) are closely related but, in fact, one can prove various rigorous implications. We mention some of them in the

context of the linear wave equation under consideration, although the same holds in a much larger class of systems including also plate and Schrödinger equations.

- Whenever exact controllability holds, solving an infinite horizon linear quadratic regulator (LQR) problem one can prove the existence of feedback operators, obtained as solutions of suitable Riccati operator equations, for which the corresponding semigroup has the property of exponential decay (see, for instance, [113]).
- Whenever the uniform exponential decay property holds, the system is exactly controllable as well, with controls supported precisely in the set where the feedback damping mechanism is active. This result is known as *Russell's stabilizability $\Rightarrow$ controllability principle* (see [150]).
- When exact controllability holds in the class of bounded control operators, stabilization holds as well (see [117,170]).

There is an extensive literature on the topic. Although in some cases, as we have mentioned, the stabilizability can be obtained as a consequence of controllability, this is not always the case. Consequently, the problem of stabilization needs often to be addressed directly and independently. The main tools for doing it are essentially the same: multipliers, microlocal analysis and Carleman inequalities.

Let us briefly mention some of the techniques and the type of results one may expect.

- The obtention of decay rates for solutions of damped wave equations has been the object of intensive research. One of the most useful tools for doing that is building new functionals, which are equivalent to the energy one, and for which differential inequalities can be obtained leading to the uniform exponential decay. These new functionals are built by perturbing the original energy one by adding terms that make explicit the effect of the mechanism on the various components of the system. At this respect it is important to note that the state of the wave equation involves in fact two components, the solution itself and its velocity. Thus, the way typical velocity feedback mechanisms affect the whole solution needs some analysis. On the other hand, in practice, the feedback is localized in part of the domain or its boundary, as controls do in controllability problems. Thus, how they affect the state everywhere else in the domain needs also some analysis. We refer to [89] where this method was introduced in the context of damped wave equations with damping everywhere in the interior of the domain. In this case the main multiplier that needs to be used is the solution of the equation itself since it allows obtaining the so-called “equipartition of energy” estimate that makes explicit the effect of the velocity feedback on the solution itself. In [103] the method was adapted to the boundary stabilization of the wave equation. In that case one needs to use the same multipliers as for the boundary observability of the wave equation with Dirichlet boundary conditions. We also refer to [171] where the method has been applied to deal with nonlinear feedback terms for which the decay is polynomial.
- In the context of nonlinear systems the obtention of uniform exponential decay rates is more delicate. Indeed, due to the presence of the nonlinearity, the exponential rate of decay may depend on the size of the solutions under consideration. In fact the nonlinearity has to satisfy some “good sign” properties at infinity to guarantee that the exponential decay rate is independent of the initial data. We refer to [170] where the uniform exponential decay has been proved for the semilinear wave equation with

linear damping concentrated on a neighborhood of the boundary. In this case the exponential decay has been directly proved by proving observability estimates without building Lyapunov functionals. We refer also to [44] where these results have been extended, by means of Strichartz estimates, to supercritical nonlinearities that energy methods do not allow to handle.

- The obtention of uniform exponential decay rates not only needs of appropriate feedback mechanisms but they also need to be supported in regions that guarantee the geometric control condition (GCC) to hold. In fact the microlocal techniques apply in stabilization problems as well (see [7]). When the GCC fails, due to the existence of Gaussian beam solutions that are exponentially concentrated away from the support of the damper, the property of uniform exponential decay fails. In that case one may only prove logarithmic decay rates for solutions with data in the domain of the operator. We refer to [18] for similar results in the context of the local energy of solutions of the wave equation in exterior domains and to [169] where this type of result has been obtained in the context of a coupled wave–heat system.
- In some cases, even if the damping mechanism is supported in a subdomain of the domain itself or of its boundary that satisfies the GCC, the uniform exponential decay may fail if the damping does not damp the energy of the system itself but rather a weaker one. This happens typically if, instead of the wave equation with velocity damping supported everywhere in the domain

$$u_{tt} - \Delta u + u_t = 0, \quad (3.35)$$

one considers

$$u_{tt} - \Delta u + K u_t = 0, \quad (3.36)$$

where $K : L^2(\Omega) \rightarrow L^2(\Omega)$ is a compact positive operator (for instance $K = (-\Delta)^{-s}$ for some $s > 0$). In the first case the energy,

$$E(t) = \frac{1}{2} \int [|u_t|^2 + |\nabla u|^2] dx,$$

satisfies the energy dissipation law

$$\frac{d}{dt} E(t) = - \int_{\Omega} |u_t|^2 dx, \quad (3.37)$$

while, in the other one, it follows that

$$\frac{d}{dt} E(t) = - \int_{\Omega} K u_t u_t dx. \quad (3.38)$$

While in the first case the energy dissipation rate is proportional to the kinetic energy, in the second one the dissipation is weaker, because of the compactness of the operator K . In the latter one cannot expect the uniform exponential decay rate to hold. In

fact, in the simplest case in which $K = (-\Delta)^{-s}$ the spectrum of the system can be computed explicitly and one sees that the spectral abscissa vanishes.

However, even if the damping is too weak, one can get a polynomial decay rate within the class of solutions in the domain of the operator. It is important to observe that dissipative semigroups are necessarily such that either the norm of the semigroup decays exponentially or the semigroup is of unit norm for all $t \geq 0$. Therefore, one may not expect any other uniform decay rate when the uniform exponential decay property fails. For that reason one needs to restrict the class of solutions under consideration. A natural way of doing that is considering initial data in the domain of the generator of the semigroup.

There are several mechanical systems in which these phenomena arise. One of them is the system of thermoelasticity in which, in several space dimensions, the damping introduced through the heat equation dissipates at most a lower-order energy with a loss of one derivative (see [110]). The same occurs often in the context of multi-structures (see for instance [86] where a system coupling two vibrating strings with a point mass is considered). We also refer to [145] where a 2- d plate with dynamical boundary conditions is considered. The polynomial decay property is proved by using a multiplier of the form $(x - x^0) \cdot \nabla u E(t)$, the novelty being that the multiplier is not linear on the solutions but rather of cubic homogeneity.

The problem of stabilization of wave equations is also intimately related to other issues in the theory of infinite-dimensional dissipative dynamical systems. We refer to [87] for an introduction to this topic. We also refer to [62] where the issue of attractors for semilinear wave equations with locally distributed damping is addressed.

4. The heat equation

4.1. Problem formulation

With the same notations as above we consider the linear controlled heat equation

$$\begin{cases} u_t - \Delta u = f \mathbb{1}_\omega & \text{in } Q, \\ u = 0 & \text{on } \Sigma, \\ u(x, 0) = u^0(x) & \text{in } \Omega. \end{cases} \quad (4.1)$$

We assume that $u^0 \in L^2(\Omega)$ and $f \in L^2(Q)$ so that (4.1) admits a unique solution

$$u \in C([0, T]; L^2(\Omega)) \cap L^2(0, T; H_0^1(\Omega)).$$

We introduce the reachable set $R(T; u^0) = \{u(T) : f \in L^2(Q)\}$. In this case the different notions of controllability can be formulated as follows:

- (a) System (4.1) is said to be *approximately controllable* if $R(T; u^0)$ is dense in $L^2(\Omega)$ for all $u^0 \in L^2(\Omega)$.
- (b) System (4.1) is *exactly controllable* if $R(T; u^0) = L^2(\Omega)$ for all $u^0 \in L^2(\Omega)$.
- (c) System (4.1) is *null controllable* if $0 \in R(T; u^0)$ for all $u^0 \in L^2(\Omega)$.

Summarizing, the following can be said about these notions.

(a) It is easy to see that exact controllability may not hold, the trivial case in which $\omega = \Omega$ being excepted.⁵ Indeed, due to the regularizing effect of the heat equation, solutions of (4.1) at time $t = T$ are smooth in $\Omega \setminus \bar{\omega}$. Therefore $R(T; u^0)$ is strictly contained in $L^2(\Omega)$ for all $u^0 \in L^2(\Omega)$.

(b) Approximate controllability holds for every open nonempty subset ω of Ω and for every $T > 0$. As we shall see, as in the case of the wave equation, the problem can be reduced to an uniqueness one that can be solved applying Holmgren's uniqueness theorem. The controls of minimal norm can be characterized as the minima of suitable quadratic functionals. As in the context of the wave equation, as a consequence of approximate controllability, we can ensure immediately that finite-approximate controllability also holds.

(c) The system being linear, null controllability implies that all the range of the semigroup generated by the heat equation is reachable too. In other words, if $0 \in R(T; u^0)$ then, $S(T)[L^2(\Omega)] \subset R(T; u^0)$, where $S = S(t)$ is the semigroup generated by the uncontrolled heat equation. This result might seem surprising in a first approach. Indeed, the sole fact that the trivial state $u^1 \equiv 0$ is reachable, implies that all the range of the semigroup is it.

But, in fact, a more careful analysis shows that the reachable set is slightly larger. We shall return to this matter.

(d) Null controllability in time T implies approximate controllability in time T . Proving it requires the use of the density of $S(T)[L^2(\Omega)]$ in $L^2(\Omega)$.

In the case of the linear heat equation this can be seen easily developing solutions in Fourier series. In the absence of control ($f \equiv 0$), the solution can be written in the form

$$u(x, t) = \sum_{j \geq 1} a_j e^{-\lambda_j t} \psi_j(x). \quad (4.2)$$

The initial datum u^0 being in $L^2(\Omega)$ is equivalent to the condition that its Fourier coefficients $\{a_j\}_{j \geq 1}$ satisfy that $\{a_j\}_{j \geq 1} \in \ell^2$.

Then, the range $S(T)[L^2(\Omega)]$ of the semigroup can be characterized as the space of functions of the form

$$\sum_{j \geq 1} a_j e^{-\lambda_j T} \psi_j(x) \quad (4.3)$$

with $\{a_j\}_{j \geq 1} \in \ell^2$. This space is small, in particular, it is smaller than any finite-order Sobolev space in Ω . But it is obviously dense in $L^2(\Omega)$ since it contains all finite linear combinations of the eigenfunctions.

If the equation contains time dependent coefficients the density of the range of the semigroup still holds, but cannot be proved by using Fourier expansions. One has rather to use a duality argument that reduces the problem to that of the backward uniqueness. This property is by now well known for the Dirichlet problem in bounded domains for

⁵In the latter controllability holds in the space H^1 instead of L^2 . However, as observed in [29], in the L^1 -setting controllability is guaranteed by means of L^1 -controls distributed everywhere in the domain.

the heat equation with lower-order terms (see [116] and [79]). It reads as follows: If $y \in C([0, T]; H_0^1(\Omega))$ solves

$$\begin{cases} y_t - \Delta y + a(x, t)y = 0 & \text{in } Q, \\ y = 0 & \text{on } \Sigma, \end{cases} \quad (4.4)$$

and

$$y(x, T) \equiv 0 \quad \text{in } \Omega,$$

then, necessarily, $y \equiv 0$.

In fact, the proof of this backward uniqueness result can be made quantitative, yielding an energy estimate which, roughly, depends exponentially on the ratio

$$R(0) = \frac{\|\nabla y(0)\|_{L^2(\Omega)}^2}{\|y(0)\|_{L^2(\Omega)}^2}$$

and the L^∞ -norm of the potential $a = a(x, t)$. Note that the initial datum $y(0)$ on that problem is assumed to be unknown but, in view of the regularity condition imposed on the solution, the ratio $R(0)$ is known to be finite. This estimate allows getting upper bounds on the L^2 -norm of solutions at time t_1 in terms of the L^2 -norm in time t_2 with $t_1 < t_2$. In particular, when $y(T) \equiv 0$ this estimate implies that $y(t) \equiv 0$ for all $0 \leq t \leq T$. In fact, one can obtain rather explicit estimates on the exponential growth of the norm of solutions backwards in time. This has been used systematically in [69] to get explicit estimates on the cost of approximate controllability.

Note also that, in fact, the density of the range of the semigroup is also true for heat equations with globally Lipschitz nonlinearities [58].

Let us now develop some of these results in more detail.

4.2. Approximate controllability

We first discuss the *approximate and the finite-approximate controllability problems*.

As we said before, system (4.1) is approximately controllable for any open, nonempty subset ω of Ω and $T > 0$. To see this one can apply Hahn–Banach’s theorem or the variational approach developed in [115] and that we have presented in the previous section in the context of the wave equation. In both cases the approximate controllability is reduced to the unique continuation property for the adjoint system

$$\begin{cases} -\varphi_t - \Delta \varphi = 0 & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x) & \text{in } \Omega. \end{cases} \quad (4.5)$$

More precisely, approximate controllability holds if and only if the following uniqueness property is true:

If φ solves (4.5) and $\varphi = 0$ in $\omega \times (0, T)$ then, necessarily, $\varphi \equiv 0$, i.e., $\varphi^0 \equiv 0$.
(4.6)

This uniqueness property holds for every open nonempty subset ω of Ω and $T > 0$ by Holmgren's uniqueness theorem.

As we have seen above in the context of the wave equation, when this unique continuation property holds, not only the system is approximately controllable but it is also finite-approximately controllable too.

Following the variational approach of [115] described in the previous subsection in the context of the wave equation, the control can be constructed as follows. First of all we observe that, the system being linear and well-posed in $L^2(\Omega)$, it is sufficient to consider the particular case $u^0 \equiv 0$. Then, for any u^1 on $L^2(\Omega)$, $\varepsilon > 0$ and E finite-dimensional subspace of $L^2(\Omega)$ we introduce the functional

$$J_\varepsilon(\varphi^0) = \frac{1}{2} \int_0^T \int_\omega \varphi^2 \, dx \, dt + \varepsilon \|(I - \pi_E)\varphi^0\|_{L^2(\Omega)} - \int_\Omega \varphi^0 u^1 \, dx, \quad (4.7)$$

where π_E denotes the orthogonal projection from $L^2(\Omega)$ over E . Strictly speaking, this functional corresponds to the case $u^0 \equiv 0$. Note however that, without loss of generality, the problem can be reduced to that particular case because of the linearity of the system under consideration. In general the functional to be considered would be

$$\begin{aligned} J_\varepsilon(\varphi^0) = & \frac{1}{2} \int_0^T \int_\omega \varphi^2 \, dx \, dt + \varepsilon \|(I - \pi_E)\varphi^0\|_{L^2(\Omega)} \\ & - \int_\Omega \varphi^0 u^1 \, dx + \int_\Omega \varphi(0) u^0 \, dx. \end{aligned} \quad (4.8)$$

The functional J_ε is continuous and convex in $L^2(\Omega)$. On the other hand, in view of the unique continuation property above, one can prove that

$$\lim_{\|\varphi^0\|_{L^2(\Omega)} \rightarrow \infty} \frac{J_\varepsilon(\varphi^0)}{\|\varphi^0\|_{L^2(\Omega)}} \geq \varepsilon. \quad (4.9)$$

Let us, for the sake of completeness, give the proof of this coercivity property.

In order to prove (4.9) let $(\varphi_j^0) \subset L^2(\Omega)$ be a sequence of initial data for the adjoint system with $\|\varphi_j^0\|_{L^2(\Omega)} \rightarrow \infty$. We normalize them by

$$\tilde{\varphi}_j^0 = \frac{\varphi_j^0}{\|\varphi_j^0\|_{L^2(\Omega)}},$$

so that $\|\tilde{\varphi}_j^0\|_{L^2(\Omega)} = 1$.

On the other hand, let $\tilde{\varphi}_j$ be the solution of (4.5) with initial data $\tilde{\varphi}_j^0$. Then

$$\begin{aligned} \frac{J_\varepsilon(\varphi_j^0)}{\|\varphi_j^0\|_{L^2(\Omega)}} &= \frac{1}{2} \|\varphi_j^0\|_{L^2(\Omega)} \int_0^T \int_\omega |\tilde{\varphi}_j|^2 \, dx \, dt \\ &\quad + \varepsilon \|(I - \pi_E)\tilde{\varphi}_j^0\|_{L^2(\Omega)} - \int_\Omega u^1 \tilde{\varphi}_j^0 \, dx. \end{aligned}$$

The following two cases may occur:

(1) $\liminf_{j \rightarrow \infty} \int_0^T \int_\omega |\tilde{\varphi}_j|^2 > 0$. In this case we obtain immediately that

$$\frac{J_\varepsilon(\varphi_j^0)}{\|\varphi_j^0\|_{L^2(\Omega)}} \rightarrow \infty.$$

(2) $\liminf_{j \rightarrow \infty} \int_0^T \int_\omega |\tilde{\varphi}_j|^2 = 0$. In this case since $\tilde{\varphi}_j^0$ is bounded in $L^2(\Omega)$, by extracting a subsequence we can guarantee that $\tilde{\varphi}_j^0 \rightharpoonup \psi^0$ weakly in $L^2(\Omega)$ and $\tilde{\varphi}_j \rightharpoonup \psi$ weakly in

$$L^2(0, T; H_0^1(\Omega)) \cap H^1(0, T; H^{-1}(\Omega)),$$

where ψ is the solution of (4.5) with initial data ψ^0 at $t = T$. Moreover, by lower semi-continuity,

$$\int_0^T \int_\omega \psi^2 \, dx \, dt \leq \liminf_{j \rightarrow \infty} \int_0^T \int_\omega |\tilde{\varphi}_j|^2 \, dx \, dt = 0$$

and therefore $\psi = 0$ in $\omega \times (0, T)$.

Holmgren's uniqueness theorem implies that $\psi \equiv 0$ in $\Omega \times (0, T)$ and consequently $\psi^0 = 0$.

Therefore, $\tilde{\varphi}_j^0 \rightharpoonup 0$ weakly in $L^2(\Omega)$ and consequently $\int_\Omega u^1 \tilde{\varphi}_j^0 \, dx$ tends to 0 as well. Furthermore, E being finite-dimensional, π_E is compact and then $\pi_E \tilde{\varphi}_j^0 \rightarrow 0$ strongly in $L^2(\Omega)$. Consequently,

$$\|(I - \pi_E)\tilde{\varphi}_j^0\|_{L^2(\Omega)} \rightarrow 1 \quad \text{as } j \rightarrow \infty.$$

Hence

$$\liminf_{j \rightarrow \infty} \frac{J_\varepsilon(\varphi_j^0)}{\|\varphi_j^0\|} \geq \liminf_{j \rightarrow \infty} \left[\varepsilon - \int_\Omega u^1 \tilde{\varphi}_j^0 \, dx \right] = \varepsilon,$$

and (4.9) follows.

Then, J_ε admits a unique minimizer $\hat{\varphi}^0$ in $L^2(\Omega)$. The control $f = \hat{\varphi}$ where $\hat{\varphi}$ solves (4.5) with $\hat{\varphi}^0$ as data is such that the solution u of (4.1) with $u^0 = 0$ satisfies

$$\|u(T) - u^1\|_{L^2(\Omega)} \leq \varepsilon, \quad \pi_E(u(T)) = \pi_E(u^1). \quad (4.10)$$

Indeed, suppose that J_ε attains its minimum value at $\hat{\varphi}^0 \in L^2(\Omega)$. Then, for any $\psi^0 \in L^2(\Omega)$ and $h \in \mathbb{R}$ we have $J_\varepsilon(\hat{\varphi}^0) \leq J_\varepsilon(\hat{\varphi}^0 + h\psi^0)$. On the other hand,

$$\begin{aligned} J_\varepsilon(\hat{\varphi}^0 + h\psi^0) &= \frac{1}{2} \int_0^T \int_\omega |\hat{\varphi} + h\psi|^2 \, dx \, dt \\ &\quad + \varepsilon \|(I - \pi_E)(\hat{\varphi}^0 + h\psi^0)\|_{L^2(\Omega)} - \int_\Omega u^1(\hat{\varphi}^0 + h\psi^0) \, dx \\ &= \frac{1}{2} \int_0^T \int_\omega |\hat{\varphi}|^2 \, dx \, dt + \frac{h^2}{2} \int_0^T \int_\omega |\psi|^2 \, dx \, dt + h \int_0^T \int_\omega \hat{\varphi} \psi \, dx \, dt \\ &\quad + \varepsilon \|(I - \pi_E)(\hat{\varphi}^0 + h\psi^0)\|_{L^2(\Omega)} - \int_\Omega u^1(\hat{\varphi}^0 + h\psi^0) \, dx. \end{aligned}$$

Thus

$$\begin{aligned} 0 &\leq \varepsilon [\|(I - \pi_E)(\hat{\varphi}^0 + h\psi^0)\|_{L^2(\Omega)} - \|(I - \pi_E)\hat{\varphi}^0\|_{L^2(\Omega)}] \\ &\quad + \frac{h^2}{2} \int_{(0,T) \times \omega} \psi^2 \, dx \, dt + h \left[\int_0^T \int_\omega \hat{\varphi} \psi \, dx \, dt - \int_\Omega u^1 \psi^0 \, dx \right]. \end{aligned}$$

Since

$$\|(I - \pi_E)(\hat{\varphi}^0 + h\psi^0)\|_{L^2(\Omega)} - \|(I - \pi_E)\hat{\varphi}^0\|_{L^2(\Omega)} \leq |h| \|(I - \pi_E)\psi^0\|_{L^2(\Omega)},$$

we obtain

$$\begin{aligned} 0 &\leq \varepsilon |h| \|(I - \pi_E)\psi^0\|_{L^2(\Omega)} + \frac{h^2}{2} \int_0^T \int_\omega \psi^2 \, dx \, dt \\ &\quad + h \int_0^T \int_\omega \hat{\varphi} \psi \, dx \, dt - h \int_\Omega u^1 \psi^0 \, dx \end{aligned}$$

for all $h \in \mathbb{R}$ and $\psi^0 \in L^2(\Omega)$.

Dividing by $h > 0$ and by passing to the limit $h \rightarrow 0$, we obtain

$$0 \leq \varepsilon \|(I - \pi_E)\psi^0\|_{L^2(\Omega)} + \int_0^T \int_\omega \hat{\varphi} \psi \, dx \, dt - \int_\Omega u^1 \psi^0 \, dx. \quad (4.11)$$

The same calculations with $h < 0$ give that

$$\left| \int_0^T \int_\omega \hat{\varphi} \psi \, dx \, dt - \int_\Omega u^1 \psi^0 \, dx \right| \leq \varepsilon \|(I - \pi_E)\psi^0\| \quad \forall \psi^0 \in L^2(\Omega). \quad (4.12)$$

On the other hand, if we take the control $f = \hat{\varphi}$ in (4.1), by multiplying in (4.1) by ψ solution of (4.5) and by integrating by parts we get that

$$\int_0^T \int_{\omega} \hat{\varphi} \psi \, dx \, dt = \int_{\Omega} u(T) \psi^0 \, dx. \quad (4.13)$$

From the last two relations it follows that

$$\left| \int_{\Omega} (u(T) - u^1) \psi^0 \, dx \right| \leq \varepsilon \|\psi^0\|_{L^2(\Omega)}, \quad \forall \psi^0 \in L^2(\Omega) \quad (4.14)$$

which is equivalent to

$$\|u(T) - u^1\|_{L^2(\Omega)} \leq \varepsilon. \quad (4.15)$$

Moreover, it also follows that

$$\int_{\Omega} (u(T) - u^1) \psi^0 \, dx = 0, \quad \forall \psi^0 \in E,$$

which shows that $\pi_E(u(T)) = \pi_E(u^1)$, and therefore (4.10) holds.

We have shown that the variational approach allows to prove the property of finite-approximate controllability, as soon as the unique continuation property for the adjoint system holds. The controls we obtain this way are those of minimal $L^2(\omega \times (0, T))$ -norm.

This method can be extended to the L^p -setting and, in particular, be used to build bang-bang controls.

4.3. Null controllability

Let us now analyze the *null controllability* problem.

This problem is also a classical one. In recent years important progresses have been done combining the variational techniques we have described and the Carleman inequalities yielding the necessary observability estimates. We shall describe some of the key ingredients of this approach in this section. However, the first results in this context were obtained in one space dimension, using the moment problem formulation and explicit estimates on the family of biorthogonal functions. We refer to [150] for a survey of the first results obtained by those techniques (see also [60,61,149]).

The null controllability problem for system (4.1) is equivalent to the following observability inequality for the adjoint system (4.5):

$$\|\varphi(0)\|_{L^2(\Omega)}^2 \leq C \int_0^T \int_{\omega} \varphi^2 \, dx \, dt \quad \forall \varphi^0 \in L^2(\Omega). \quad (4.16)$$

Due to the time-irreversibility of the system (4.5), (4.16) is not easy to prove. For instance, multiplier methods, that are so efficient for wave-like equations, do not apply.

Nevertheless, inequality (4.16) is rather weak. Indeed, in contrast with the situation we encountered when analyzing the wave equation, in the present case, getting estimates on the solution of the adjoint system at $t = 0$ is much weaker than getting estimates of φ^0 at time $t = T$. Indeed, due to the very strong irreversibility of the adjoint system (4.5), which is well posed in the backward sense of time, one cannot get estimates of the initial datum φ^0 out of the estimate (4.16). We shall come back to this issue later.

There is an extensive literature on the null control of the heat equation. In [150] the boundary null controllability of the heat equation was proved in one space dimension using moment problems and classical results on the linear independence in $L^2(0, T)$ of families of real exponentials. On the other hand, in [149] it was shown that *if the wave equation is exactly controllable for some $T > 0$ with controls supported in ω , then the heat equation (4.1) is null controllable for all $T > 0$ with controls supported in ω* . We refer to [6] for a systematic and more recent presentation of this method. As a consequence of this result and in view of the controllability results of the previous section for the wave equation, it follows that the heat equation (4.1) is null controllable for all $T > 0$ provided ω satisfies the GCC and the observability inequality (4.16) holds.

The fact that the control time $T > 0$ is arbitrary for the heat equation is in agreement with the intrinsic infinite velocity of propagation of the heat model. However, the GCC does not seem to be a natural sharp condition in the context of the heat equation. Indeed, in view of the diffusion and regularizing process that the heat equation induces one could expect the heat equation to be null-controllable from any open nonempty subset ω . This result was proved by Lebeau and Robbiano [108]. Simultaneously, the same was proved independently by Imanuvilov in [94] (see also [76]) for a much larger class of heat equations with lower-order potentials by using parabolic Carleman inequalities. We shall return to this issue later.

Let us first discuss the method in [108] which is based on the Fourier decomposition of solutions. A simplified presentation was given in [109] where the linear system of thermoelasticity was also addressed. The main ingredient of it is the following observability estimate for the eigenfunctions of the Laplace operator.

THEOREM 4.1 [108,109]. *Let Ω be a bounded domain of class C^∞ . For any nonempty open subset ω of Ω there exist positive constants $B, C > 0$ such that*

$$C e^{-B\sqrt{\mu}} \sum_{\lambda_j \leq \mu} |a_j|^2 \leq \int_{\omega} \left| \sum_{\lambda_j \leq \mu} a_j \psi_j(x) \right|^2 dx \quad (4.17)$$

for all $\{a_j\} \in \ell^2$ and for all $\mu > 0$.

The proof of (4.17) is based on Carleman inequalities (see [108] and [109]).

Although the constant in (4.17) degenerates exponentially as $\mu \rightarrow \infty$, it is important that it does it exponentially on $\sqrt{\mu}$ and not exponentially on μ or any other larger power of μ . As we shall see, the strong dissipativity of the heat equation allows compensating this fact. Indeed, estimate (4.17) provides a measure of the degree of linear independence of the traces of linear finite combinations of eigenfunctions over ω . By inspection of the

Gaussian heat kernel it can be shown that this estimate, i.e., the degeneracy of the constant in (4.17) as $\exp(-B\sqrt{\mu})$ for some $B > 0$, is sharp even in $1-d$.

As a consequence of (4.17) one can prove that the observability inequality (4.16) holds for solutions of (4.5) with initial data in $E_\mu = \text{span}\{\psi_j\}_{\lambda_j \leq \mu}$, the constant being of the order of $\exp(B\sqrt{\mu})$. This shows that the projection of solutions of (4.1) over E_μ can be controlled to zero with a control of size $\exp(B\sqrt{\mu})$.⁶ Thus, when controlling the frequencies $\lambda_j \leq \mu$ one increases the $L^2(\Omega)$ -norm of the high frequencies $\lambda_j > \mu$ by a multiplicative factor of the order of $\exp(B\sqrt{\mu})$. However, solutions of the heat equation (4.1) without control ($f = 0$) and such that the projection of the initial data over E_μ vanishes, decay in $L^2(\Omega)$ at a rate of the order of $\exp(-\mu t)$. This can be easily seen by means of the Fourier series decomposition of the solution. Thus, if we divide the time interval $[0, T]$ in two parts $[0, T/2]$ and $[T/2, T]$, we control to zero the frequencies $\lambda_j \leq \mu$ in the interval $[0, T/2]$ and then allow the equation to evolve without control in the interval $[T/2, T]$, it follows that, at time $t = T$, the projection of the solution u over E_μ vanishes and the norm of the high frequencies does not exceed the norm of the initial data u^0 .

This argument allows to control to zero the projection over E_μ for any $\mu > 0$ but not the whole solution. To do that an iterative method is needed in which the interval $[0, T]$ has to be decomposed in a suitably chosen sequence of subintervals $[T_k, T_{k+1})$ and the argument above is applied in each subinterval to control an increasing range of frequencies $\lambda_j \leq \mu_k$ with $\mu_k \rightarrow \infty$ at a suitable rate. We refer to [108] and [109] for the proof.

Once (4.16) is known to hold for the solutions of the adjoint heat equation (4.5) one can obtain the control with minimal $L^2(\omega \times (0, T))$ -norm among the admissible ones. To do that it is sufficient to minimize the functional

$$J(\varphi^0) = \frac{1}{2} \int_0^T \int_\omega \varphi^2 \, dx \, dt + \int_\Omega \varphi(0) u^0 \, dx \quad (4.18)$$

over the Hilbert space

$$H = \left\{ \varphi^0: \text{ the solution } \varphi \text{ of (4.5) satisfies } \int_0^T \int_\omega \varphi^2 \, dx \, dt < \infty \right\},$$

endowed with its canonical norm.

To be more precise, H should be defined as the completion of $\mathcal{D}(\Omega)$ with respect to the norm

$$\|\varphi^0\|_H = \left(\int_0^T \int_\omega \varphi^2 \, dx \, dt \right)^{1/2}.$$

The space H is very large. In fact, due to the regularizing effect of the heat equation, any initial (at time $t = T$) datum φ^0 of the adjoint heat equation in H^{-s} , whatever $s > 0$ is,

⁶In fact the same is true for any evolution equation allowing a Fourier expansion on the basis of these eigenfunctions (Schrödinger, plate, wave equations, etc.). The novelty of the argument in [108] when applied to the heat equation is that its dissipative effect is able to compensate the growth of the control as μ tends to infinity, a fact that does not hold for conservative systems. This is sharp and natural, to some extent, since we know that the wave equation does not have the property of being controllable from any open set.

belongs to H because the corresponding solutions satisfy $\varphi \in L^2(\omega \times (0, T))$. We shall return in the following section to the discussion of the nature and structure of this space.

Observe that J is convex and continuous in H . On the other hand, (4.16) guarantees the coercivity of J and the existence of its minimizer. The minimizer of J provides the control we are looking for, which is of minimal $L^2(\omega \times (0, T))$ -norm.

There is an easy way to build null controls and avoiding working in the space H . Indeed, we can build, for all $\varepsilon > 0$, an approximate control f_ε such that the solution u_ε of (4.1) satisfies the condition

$$\|u_\varepsilon(T)\|_{L^2(\Omega)} \leq \varepsilon. \quad (4.19)$$

Recall that, for this to be true, the unique continuation property (4.6) of the adjoint system suffices. But, the fact that the observability inequality (4.16) holds adds an important information to this: the sequence of controls $\{f_\varepsilon\}_{\varepsilon>0}$ is uniformly bounded. Assuming for the moment that this holds let us conclude the null controllability of (4.1) out of these results. In view of the fact that controls $\{f_\varepsilon\}_{\varepsilon>0}$ are uniformly bounded in $L^2(\omega \times (0, T))$, by extracting subsequences, we have $f_\varepsilon \rightarrow f$ weakly in $L^2(\omega \times (0, T))$. Using the continuous dependence of the solutions of the heat equation (4.1) on the right-hand side term, we can show that $u_\varepsilon(T)$ converges to $u(T)$ weakly in $L^2(\Omega)$. In view of (4.19) this implies that $u(T) \equiv 0$. The limit control f then fulfills the null-controllability requirement.

In order to see that the controls f_ε are bounded, we have to use its structure. Note that $f_\varepsilon = \hat{\varphi}_\varepsilon$, where $\hat{\varphi}_\varepsilon$ solves (4.5) with initial data at time $t = T$ obtained by minimizing the functional (4.8) when $E = \{0\}$ and $u^1 \equiv 0$. At the minimizer $\hat{\varphi}_\varepsilon^0$ we have $J_\varepsilon(\varphi_\varepsilon^0) \leq J_\varepsilon(0) = 0$. This implies that

$$\int_{\omega \times (0, T)} |\hat{\varphi}_\varepsilon|^2 dx dt \leq \|u^0\|_{L^2(\Omega)} \|\varphi_\varepsilon(0)\|_{L^2(\Omega)}.$$

This, together with the observability inequality (4.16), implies that

$$\left[\int_{\omega \times (0, T)} |f_\varepsilon|^2 dx dt \right]^{1/2} = \left[\int_{\omega \times (0, T)} |\hat{\varphi}_\varepsilon|^2 dx dt \right]^{1/2} \leq C \|u^0\|_{L^2(\Omega)},$$

which yields the desired bound on the approximate controls.

Note that, the approximate controllability in itself (or, in other words, the unique continuation property of the adjoint system) does not yield this bound. We have rather used the fact that observability inequality (4.16) holds. The argument above simply avoids minimizing the functional in H , a space whose nature will be investigated later.

As a consequence of the internal null controllability property of the heat equation one can deduce easily the null boundary controllability with controls in an arbitrarily small open subset of the boundary. To see this it is sufficient to extend the domain Ω by a little open subset attached to the subset of the boundary where the control needs to be supported. The arguments above allow to control the system in the large domain by means of a control supported in this small added domain. The restriction of the solution to the original domain

satisfies all the requirements and its restriction or trace to the subset of the boundary where the control had to be supported, provides the control we were looking for.

Note however that the boundary control problem may be addressed directly. As a consequence of Holmgren's uniqueness theorem, the corresponding unique continuation result holds and, as a consequence we obtain approximate and finite-approximate controllability. On the other hand, Carleman inequalities yield the necessary observability inequalities to derive null controllability as well (see for instance [64,76,94]). As a consequence of this, as in the context of the internal control problem, null controllability holds in an arbitrarily small time and with boundary controls supported in an arbitrarily small open nonempty subset of the boundary.

The method of proof of the null controllability property we have described is based on the possibility of expanding solutions in Fourier series. Thus it can be applied in a more general class of heat equations with variable but time-independent coefficients. The same can be said about the methods of [149]. In the following section we shall present a direct Carleman inequality approach proposed and developed in [94] and [76] for the parabolic problem which allows circumventing this difficulty.

Recently Miller in [131] used for control a transformation inspired in the so-called Kananai transform, previously used by Phung [140] to analyze the cost of controllability for Schrödinger equations, allowing to write the solutions of the heat equation and their controls in terms of those of the wave equation, to derive null controllability results for the heat equation as a consequence of the exact controllability of the wave equation. This approach, the so-called control transmutation method according to the terminology in [132], plays the role in the physical space of that used by Russell [149] which consists in performing a change of variable in the frequency domain. Both approaches give similar results. The advantage of the transmutation method is that it allows getting explicit estimates of the norms of the controls more easily and that it does not require of any eigenfunction decomposition of solutions. Its drawback, as in Russell's approach, is that it only applies to heat equations with coefficients which are independent of time.

In the last section devoted to open problems we shall return to these issues. Indeed, many interesting and deep questions remain open in this field, related to the connections between the geometry of the domains Ω and ω under consideration and the best constants in observability inequalities, in particular with the best constant B in (4.17).

4.4. Parabolic equations of fractional order

The iterative argument developed in [108] and [109] based on the spectral estimate (4.17) suggests that the regularizing effect of the heat equation is far too much to guarantee the null controllability. Indeed, controlling low frequencies $\lambda_j \leq \mu$ costs $\exp(C\sqrt{\mu})$ while the dissipation rate of the controlled one is $\exp(-\mu)$. In view of this it would be natural to consider equations of the form

$$u_t + (-\Delta)^\alpha u = f \mathbb{1}_\omega, \quad (4.20)$$

where $(-\Delta)^\alpha$ denotes the α -th power of the Dirichlet Laplacian. This problem was addressed in [128] where it was proved that:

- The system is null controllable for all $\alpha > 1/2$. This result is not hard to guess from the iterative construction above. The range $\alpha > 1/2$ is that in which the dissipative effect dominates and compensates the increasing cost of controlling the low frequencies, as its range increases. We refer also to [134] where this problem has been further investigated providing, in particular, an estimate on the cost of controllability when $\alpha > 1/2$.
- The system is not null controllable when $\alpha \leq 1/2$ when the input space is one dimensional (*lumped control*). In particular, null controllability fails in the critical case $\alpha = 1/2$. This is due to the following result on the lack of linear independence of the sums of real exponentials that was previously proved in [126]:

PROPOSITION 4.1. Assume that $\alpha \leq 1/2$. Then there is no sequence $\{\rho_n\}_{n \geq 1}$ of positive weights, i.e., $\rho_n > 0$ for all $n \geq 1$, such that

$$\sum_{n \geq 1} \rho_n |a_n|^2 \leq \int_0^T \left| \sum_{n \geq 1} a_n e^{-n^{2\alpha} t} \right|^2 dt \quad (4.21)$$

for all sequence of real numbers $\{a_n\}$.

The inequality (4.21) is the one that is required to obtain an observability inequality of the form (4.16) for the solutions of the adjoint fractional parabolic equation

$$-\varphi_t + (-\Delta)^\alpha \varphi = 0, \quad (4.22)$$

in one space dimension.

Note that inequalities of the form (4.21) are well known to hold when $\alpha > 1/2$ (see [60]).

The fact that the inequality does not hold, whatever the weights $\{\rho_n\}_{n \geq 1}$ are, indicates that not only (4.16) does not hold but that any weakened version of it fails as well. The lack of controllability of the system for $\alpha \leq 1/2$ has then some catastrophic nature in the sense that it cannot be compensated by restricting the class of initial data under consideration.

The fact that inequalities (4.21) fail to hold was proved in [126] in the context of the control of the heat equation in unbounded domains. There it was proved that, despite of the fact that the model has infinite speed of propagation, there is no compactly supported smooth initial data that can be controlled to zero by means of L^2 -controls localized in a bounded set. This result was later interpreted (and extended in a significant way) in [55] as a backward unique continuation one, in the absence of boundary conditions in the complement of a bounded set. The proof in [126] was based on the fact that, when writing the heat equation in conical domains with similarity variables, the underlying elliptic operator turns out to have a discrete spectrum and the eigenvalues grow in a linear way. This corresponds precisely to the critical case $\alpha = 1/2$ in model (4.22) in one space dimension in which, according to Proposition 4.1, controllability fails to hold.

4.5. Carleman inequalities for heat equations with potentials

The null controllability of the heat equation with lower-order time-dependent terms of the form

$$\begin{cases} u_t - \Delta u + a(x, t)u = f \mathbb{1}_\omega & \text{in } Q, \\ u = 0 & \text{on } \Sigma, \\ u(x, 0) = u^0(x) & \text{in } \Omega, \end{cases} \quad (4.23)$$

has been proved for the first time in a series of works by Fursikov and Imanuvilov (see for instance [28, 75, 76, 94, 98] and the references therein). Their approach is based on a direct application of Carleman inequalities to the adjoint system

$$\begin{cases} -\varphi_t - \Delta \varphi + a(x, t)\varphi = 0 & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x) & \text{in } \Omega. \end{cases} \quad (4.24)$$

More precisely, observability inequalities of the form (4.16) are directly derived for system (4.24), for all $T > 0$ and any open nonempty subset ω of Ω . We shall describe this approach in the following section.

This method has been extremely successful when dealing with observability inequalities for parabolic problems as the reader may see from the articles in the list of references in the end of this paper devoted to that issue. The method being very flexible, the same ideas have been applied in a variety of problems, including Navier–Stokes equations, and also heat equations with variable coefficients in the principal part. Carleman inequalities require these coefficients to be, roughly, Lipschitz continuous (the latter can be weakened to a suitable L^p bound on its derivatives). But, as far as we know, there is no result in the literature showing the lack of null controllability of the heat equation with bounded measurable coefficients. This is an interesting and possibly difficult open problem that we shall discuss later on in more detail.

4.6. Bang-bang controls

A slight change on the functional J_ε introduced in (4.7) to prove finite-approximate controllability allows building *bang-bang* controls. Indeed, we set

$$J_{\text{bb},\varepsilon}(\varphi^0) = \frac{1}{2} \left(\int_0^T \int_\omega |\varphi| \, dx \, dt \right)^2 + \varepsilon \|(I - \pi_E)\varphi^0\|_{L^2(\Omega)}^2 - \int_\Omega u^1 \varphi^0 \, dx. \quad (4.25)$$

The functional $J_{\text{bb},\varepsilon}$ is continuous and convex in $L^2(\Omega)$ and satisfies the coercivity property (4.9) too.

Let $\hat{\varphi}^0$ be the minimizer of $J_{\text{bb},\varepsilon}$ in $L^2(\Omega)$ and $\hat{\varphi}$ the corresponding solution of (4.5). We set

$$f = \int_0^T \int_{\omega} |\hat{\varphi}| \, dx \, dt \, \text{sgn}(\hat{\varphi}), \quad (4.26)$$

where sgn is the multivalued sign function: $\text{sgn}(s) = 1$ if $s > 0$, $\text{sgn}(0) = [-1, 1]$ and $\text{sgn}(s) = -1$ when $s < 0$. The control f given in (4.26) is such that the solution u of (4.1) with null initial data satisfies (4.10).

Due to the regularizing effect of the heat equation, the solution $\hat{\varphi}$ is analytic and its zero set is of zero $(n+1)$ -dimensional Lebesgue measure. Thus, the control f in (4.26) is of bang-bang form, i.e., $f = \pm\lambda$ a.e. in $\omega \times (0, T)$ where $\lambda = \int_0^T \int_{\omega} |\hat{\varphi}| \, dx \, dt$.

We have proved the following result:

THEOREM 4.2 [175]. *Let ω be any open nonempty subset of Ω and $T > 0$ be any positive control time. Then, for any $u^0, u^1 \in L^2(\Omega)$, $\varepsilon > 0$ and finite-dimensional subspace E of $L^2(\Omega)$, there exists a bang-bang control $f \in L^\infty(Q)$ such that the solution u of (4.1) satisfies (4.10). The control that the variational approach provides is of minimal L^∞ -norm among the class of admissible ones.*

The fact that the control obtained when minimizing $J_{\text{bb},\varepsilon}$ is of minimal L^∞ -norm was proved in [56] by using a classical duality principle (see [54]). In [56] we also considered linear equations with potentials and semilinear ones. In those cases the fact that the control obtained this way is bang-bang is less clear since one cannot use the analyticity of solutions to directly obtain that the zero-set of solutions of the adjoint system is of null Lebesgue measure. In that case one rather needs to use more sophisticated results in that direction as those in [3]. In general, the method described above always yields quasi bang-bang controls, as for the wave equation. Whether controls are actually of bang-bang form is a widely open problem. Obviously, it can be viewed as a problem of unique continuation: *Does the fact that the measure of the zero set $\{(x, t): \varphi(x, t) = 0\}$ be positive, imply that $\varphi \equiv 0$?*

Bang-bang controls also exist in the context of null controllability. This is specific to the heat equation and, as we shall see, its very strong dissipative effect plays a key role on the proof of this result.

To address the problem of null controllability we have to take $\varepsilon = 0$ and consider the functional

$$J_{\text{bb}}(\varphi^0) = \frac{1}{2} \left(\int_0^T \int_{\omega} |\varphi| \, dx \, dt \right)^2 - \int_{\Omega} u^1 \varphi^0 \, dx. \quad (4.27)$$

The functional J_{bb} is well defined and continuous in the Banach space X constituted by the solutions of the adjoint heat equation (4.5) such that $\varphi \in L^1(\omega \times (0, T))$. The space X is endowed with its canonical norm, namely, $\|\varphi\|_X = \int_0^T \int_{\omega} |\varphi| \, dx \, dt$. Note that the coercivity

of J_{bb} in X is not obvious at all. In fact, for that to be true, one has to show an observability inequality of the form

$$\|\varphi(0)\|_{L^2(\Omega)} \leq C \int_0^T \int_{\omega} |\varphi| \, dx \, dt. \quad (4.28)$$

Note that we have not mentioned this inequality so far. It is an observability inequality in which the localized observation is made in the L^1 -norm instead of the L^2 -norm considered so far. This estimate was proved in Proposition 3.2 of [70] in order to get bounded controls using the regularizing effect of the heat equation. In fact it was proved not only for the heat equation (4.5) but also for equations with zeroth-order potentials with explicit estimates on the observability constant in terms of the potential.

According to the observability estimate (4.28) the functional J_{bb} is also coercive.

However, because of the lack of reflexivity of the space X , the existence of the minimizer of J_{bb} in X may not be guaranteed. Indeed, the minimizer can only be guaranteed to exist in the closure of X with respect to the weak convergence in the sense of measures. We denote that space by $\tilde{X}$. More precisely, $\tilde{X}$ is the space of solutions of the adjoint heat equation whose restriction to $\omega \times (0, T)$ is a bounded measure. We denote by $\hat{\varphi}$ the minimizer of J_{bb} in $\tilde{X}$.

We claim that the minimizer is in fact smooth. This is due to the regularizing effect and to an improved observability estimate. Indeed, in view of the results in [70] we can improve (4.28) to obtain

$$\int_0^{T-\tau} \int_{\Omega} |\varphi|^2 \, dx \, dt \leq C_{\tau} \left[\int_0^T \int_{\omega} |\varphi| \, dx \, dt \right]^2. \quad (4.29)$$

This estimate can be extended by density to $\tilde{X}$. In that case the L^1 -norm on $\omega \times (0, T)$ has to be replaced by the total measure of the solution on that set.

Estimate (4.29) shows, in particular, that the minimizer $\hat{\varphi}$ is such that

$$\hat{\varphi} \in L^2(\Omega \times (0, T - \tau)) \quad \forall \tau > 0.$$

Thus, the minimizer is a smooth solution of the heat equation except possibly at $t = T$.

Accordingly, the Euler–Lagrange equations associated to the minimizer show the existence of a null control for the heat equation of the form

$$f \in \int_0^T \int_{\omega} |\hat{\varphi}| \, dx \, dt \operatorname{sgn} \hat{\varphi}.$$

This control is of quasibang-bang form. In fact it is strictly bang-bang since, because of the analyticity of solutions of the heat equation, its zero set is of null Lebesgue measure.

4.7. Discussion and comparison

In these sections we have presented the main controllability properties under consideration and some of the fundamental results both for the wave and the heat equation. There are some clear differences in what concerns the way each of these equations behaves:

- For the wave equation exact and null controllability are equivalent notions. However, for the heat equation, the exact controllability property may not hold and the null controllability property is the most natural one to address. Null controllability is in fact equivalent to an apparently stronger property of control to trajectories, which guarantees that every state which is the value at the final time of a solution of the uncontrolled equation is reachable from any initial datum by means of a suitable control.
- The exact controllability property for the wave equation holds provided the geometric control condition is satisfied. This imposes severe restrictions on the subset where the control acts and requires the control time to be large enough. At the contrary, the null controllability property for the heat equation holds in an arbitrarily small time and with controls in any open nonempty subset of the domain.

In view of this, we can say that, although exact controllability may not hold for the heat equation, at the level of null controllability it behaves much better than the wave equation since no geometric requirements are needed for it to hold in an arbitrarily small time.

We have described above a number of methods allowing to prove that the heat equation is null controllable for all $T > 0$ whenever the corresponding wave equation is controllable for some time $T > 0$. This result is not optimal when applied to the constant coefficient heat equation (in particular since geometric restrictions are needed on the subset where the control applies) but it has the advantage of yielding results in situations in which the parabolic methods described are hard to apply directly. This procedure has been recently used in a number of situations:

- In [71] the problem of the null controllability of the 1- d heat equation with variable coefficients in the principal part is addressed. Using this procedure, null controllability is proved in the class of BV coefficients. More recently, this result has been improved in [1] to equations with bounded measurable coefficients. The method in [1] consists in extending the statement in Theorem 4.1 to 1- d Sturm–Liouville problems with bounded measurable coefficients, using the theory of quasiregular mappings.

However, the global Carleman inequalities do not seem to yield this result, since globally Lipschitz coefficients are required. In [51] parabolic equations with piecewise constant coefficients were addressed by means of global Carleman inequalities in the multidimensional case. There observability and null controllability was proved but only under suitable monotonicity conditions on the coefficients along interfaces.

This is due to the fact that, when applying Carleman inequalities to heat equations with discontinuous coefficients, integration by parts generates some singular terms on the interfaces. These terms cannot be absorbed as lower-order ones. Therefore, a sign condition has to be imposed. This type of monotonicity condition is natural in the context of wave equations where it is known that interfaces may produce trapped waves [123]. But as far as we know, there is no evidence for the need of this kind of monotonicity condition for parabolic equations. This is an interesting and probably deep open problem. We shall return to this matter below.

Of course, in 1- d this monotonicity conditions are unnecessary, as the results in [1] and [71] show. But the situation is unclear in the multidimensional case.

- Recently an unified approach for Carleman inequalities for parabolic and hyperbolic equations has been presented in [72]. There a pointwise weighted identity is derived for general second-order operators. In this way the author is able to recover the existing Carleman inequalities for heat, wave, plate and Schrödinger equations. A viscous Schrödinger equation in between the heat and the Schrödinger equations is also addressed.
- In [41] the heat equation in a planar 1- d network is addressed. A number of null controllability results are proved by means of this procedure, as a consequence of the previously proved ones for the corresponding wave equation in the same network. In the context of networks, the wave equation is easier to deal with since one may use propagation arguments, sidewise energy estimates, D'Alembert's formula, ... So far the null controllability of parabolic equations on networks has not been addressed directly by means of Carleman inequalities. The difficulty for doing that is the treatment of the nodes of the network where various segments are interconnected. There, as in the context of parabolic equations with discontinuous coefficients, it is hard to match the Carleman inequality along each segment and to deduce a global observability estimate. On the other hand, as the spectral analysis shows, depending of the structure of the network and the mutual lengths of the segments entering on it, there may exist concentrated eigenfunctions making observability impossible. Thus, the difficulty is not merely technical. The understanding of this issue by means of Carleman inequalities is an interesting open subject.

As we mentioned above, more recently similar results have been obtained transforming the controls of the wave equation into controls for the heat one [132,140]. Both approaches are limited to the case where the coefficients of the equations are independent of time. As far as we know, there is no systematic method to transfer control results for wave equations into control results for heat equations with potentials depending on x and t .

We have also observed important differences in what concerns bang-bang controls. Bang-bang controls exist for the heat equation both in the context of approximate and exact controllability. However, for the wave equation, only quasibang-bang controls can be found and this in the framework of approximate controllability. The same analysis fails for the exact controllability of the wave equation.

5. Sharp observability estimates for the linear heat equation

5.1. Sharp estimates

In the previous section we have mentioned that the heat equation (4.23) with lower-order potentials depending both on x and t is null controllable. This is equivalent to an observability inequality for the adjoint heat equation (4.24). The only existing method that allows dealing with equations of this form are the so-called global Carleman inequalities. They were introduced in this context by Imanuvilov (see [94], and the books [74,76]) and have allowed to solve a significant amount of complex control problems for parabolic equations,

including the Navier–Stokes equations (see [36], [68,77], for instance). In this section we present the inequality in the form it was derived in [69] following the method in [76]. The observability constant depends on the norm of the potential in an, apparently, unexpected manner. But, as we shall see in the following section, according to the recent results in [53], the estimate turns out to be sharp.

The following holds.

PROPOSITION 5.1 [76,69]. *There exists a constant $C > 0$ that only depends on Ω and ω such that the following inequality holds*

$$\|\varphi(0)\|_{L^2(\Omega)}^2 \leq \exp\left[C\left(1 + \frac{1}{T} + T\|a\|_\infty + \|a\|_\infty^{2/3}\right)\right] \int_0^T \int_\omega \varphi^2 \, dx \, dt \quad (5.1)$$

for any φ solution of (4.24), for any $T > 0$ and any potential $a \in L^\infty(Q)$.

Furthermore, the following global estimate holds

$$\int_0^T \int_\Omega e^{-A(1+T)/(T-t)} \varphi^2 \, dx \, dt \leq \exp\left[C\left(1 + \frac{1}{T} + \|a\|_\infty^{2/3}\right)\right] \int_0^T \int_\omega \varphi^2 \, dx \, dt \quad (5.2)$$

with a constant A that only depends on the domains Ω and ω as well.

REMARK 5.1. Several remarks are in order.

1. Note that (5.1) provides the observability inequality for the adjoint heat equation (4.24) with an explicit estimate of the observability constant, depending on the control time T and the potential a . The observability inequality (5.2) differs from that in (5.1), on the fact that it provides a global estimate on the solution in $\Omega \times (0, T)$, but with a weight function that degenerates exponentially at $t = T$. In fact, using Carleman inequalities one first derives (5.2) to later obtain the pointwise (in time $t = 0$) estimate out of it. When doing that one needs to apply Gronwall's inequality for the time evolution of the $L^2(\Omega)$ norm of the solution. This yields the extra term $e^{T\|a\|_\infty}$ in the observability constant.

2. Inequality (5.1) plays an important role when dealing, for instance, with the null control of nonlinear problems. Using this explicit observability estimate and, in particular, the fact that it depends exponentially on the power $2/3$ of the potential a , in [70] the null controllability was proved for a class of semilinear heat equations with nonlinearities growing at infinity slower than $s \log^{3/2}(s)$. This is a surprising result since, in this range of nonlinearities, in the absence of control, solutions may blow up in finite time. The presence of the control avoids blow-up to occur and makes the solution reach the equilibrium at time $t = T$.

The estimates in Proposition 5.1 are a direct consequence of the Carleman estimates that we briefly describe now.

We introduce a function $\eta^0 = \eta^0(x)$ such that

$$\begin{cases} \eta^0 \in C^2(\overline{\Omega}), \\ \eta^0 > 0 & \text{in } \Omega, \eta^0 = 0 \text{ in } \partial\Omega, \\ \nabla \eta^0 \neq 0 & \text{in } \overline{\Omega} \setminus \omega. \end{cases} \quad (5.3)$$

The existence of this function was proved in [76]. In some particular cases, for instance when Ω is star shaped with respect to a point in ω , it can be built explicitly without difficulty. But the existence of this function is less obvious in general, when the domain has holes or its boundary oscillates, for instance.

Let $k > 0$ be such that

$$k \geq 5 \max_{\overline{\Omega}} \eta^0 - 6 \min_{\overline{\Omega}} \eta^0$$

and let

$$\beta^0 = \eta^0 + k, \quad \bar{\beta} = \frac{5}{4} \max \beta^0, \quad \rho^1(x) = e^{\lambda \bar{\beta}} - e^{\lambda \beta^0}$$

with λ sufficiently large. Let be finally

$$\gamma = \frac{\rho^1(x)}{t(T-t)}; \quad \rho(x, t) = \exp(\gamma(x, t))$$

and the space of functions

$$Z = \{q : Q \rightarrow \mathbb{R} : q \in C^2(\overline{Q}), q = 0 \text{ in } \Sigma\}.$$

The following Carleman inequality holds.

PROPOSITION 5.2 [76]. *There exist positive constants $C_*, s_1 > 0$ such that*

$$\begin{aligned} & \frac{1}{s} \int_Q \rho^{-2s} t (T-t) [|q_t|^2 + |\Delta q|^2] dx dt \\ & + s \int_Q \rho^{-2s} t^{-1} (T-t)^{-1} | \nabla q |^2 dx dt + s^3 \int_Q \rho^{-2s} t^{-3} (T-t)^{-3} q^2 dx dt \\ & \leq C_* \left[\int_Q \rho^{-2s} | \partial_t q + \Delta q |^2 dx dt \right. \\ & \quad \left. + s^3 \int_0^T \int_\omega \rho^{-2s} t^{-3} (T-t)^{-3} q^2 dx dt \right] \end{aligned} \quad (5.4)$$

for all $q \in Z$ and $s \geq s_1$.

Moreover, C_* depends only on Ω and ω and s_1 is of the form

$$s_1 = s_0(\Omega, \omega)(T + T^2),$$

where $s_0(\Omega, \omega)$ only depends on Ω and ω .

We refer to the Appendix in [69] for a proof of this Carleman inequality.

Applying (5.4) with $q = \varphi$ we deduce easily (5.2), taking into account that the first term on the right-hand side of (5.4) coincides with $a\varphi$ when φ is a solution of (4.24). In order to absorb this term we make use of the third term on the left-hand side of (5.4). This imposes the choice of the parameter s as being of the order of $\|a\|_\infty^{2/3}$ and yields that factor on the exponential observability constants in (5.1) and (5.2).

As observed in [69], when $a \equiv 0$ in the adjoint heat equation (4.24) or, more generally, when the potential is independent of t , these estimates can be written in terms of the Fourier coefficients $\{a_k\}$ of the datum of the solution of the adjoint system at $t = T$,

$$\varphi^0(x) = \sum_{k \geq 1} a_k \psi_k(x).$$

The following holds.

THEOREM 5.1 [69]. *Let $T > 0$ and ω be an open nonempty subset of Ω . Then, there exist $C, c > 0$ such that*

$$\sum_{k=1}^{\infty} |a_k|^2 e^{-c\sqrt{\lambda_k}} \leq C \int_0^T \int_{\omega} \varphi^2 \, dx \, dt \quad (5.5)$$

for all solution of (4.5), where $\{\psi_k\}$ denotes the orthonormal basis of $L^2(\Omega)$ constituted by the eigenfunctions of the Dirichlet Laplacian, $\{\lambda_k\}$ the sequence of corresponding eigenvalues and $\{a_k\}$ the Fourier coefficients of φ^0 on this basis.

REMARK 5.2. Note that the left-hand side of (5.5) defines a norm of φ^0 that corresponds to the one in the domain of the operator $\exp(-c\sqrt{-\Delta})$. Characterizing the best constant c in (5.5) in terms of the geometric properties of the domains Ω and ω is an open problem. Obviously, the constant may also depend on the length of the time interval T . The problem may be made independent of T by considering the analogue in infinite time,

$$\sum_{k=1}^{\infty} |a_k|^2 e^{-c\sqrt{\lambda_k}} \leq C \int_{-\infty}^0 \int_{\omega} \varphi^2 \, dx \, dt, \quad (5.6)$$

φ being now the solution of the adjoint system for $t \leq 0$.

As far as we know the characterization of the best constant $c > 0$ in (5.6) is an open problem. This problem is intimately related to the characterization of the best constant $A > 0$ in (5.2) for $a \equiv 0$, which is also an open problem.

By inspection of the proof of the inequality (5.2), as a consequence of the Carleman inequality (5.4), one sees that A depends on the geometric properties of the weight η^0 in (5.3). But how this is translated into the properties of the domains Ω and ω is to be investigated. Some lower bounds on A have been obtained in terms of the Gaussian heat kernels in [69] and [131]. But further investigation is needed for a complete characterization of the sharp value of A as well.

Observe also that the observability inequality (5.2) is stronger than (5.1). Indeed, (5.2) provides a global estimate on φ away from $t = 0$ and yields, in particular, (5.6) with weights $e^{-c\sqrt{\lambda_k}}$. Inequality (5.1) is much weaker since it provides only estimates on $\varphi(0)$ and therefore involves weights of the form $e^{-\lambda_k T}$ in its Fourier representation.

REMARK 5.3. In [69] these estimates were used to obtain sharp estimates on the cost of approximate and finite-approximate controllability, i.e., on the size of the control f_ε needed to reach (4.10). As we mentioned above, roughly speaking, when the final datum is not reachable, for instance when u^1 is the characteristic function of some measurable subset of Ω , the cost of controlling to an ε distance grows exponentially as ε tends to zero.

In the same article the connections between optimal control and approximate control were also explored and quantified. It is well known that the approximate controllability property can be achieved as a limit of optimal control problems with a penalization parameter k tending to ∞ that enhances the requirement of getting close to the target. More precisely, when looking for the optimal control $f \in L^2(\omega \times (0, T))$ that minimizes the functional

$$I_k(f) = \frac{1}{2} \int_0^T \int_\omega f^2 \, dx \, dt + \frac{k}{2} \|u(T) - u^1\|_{L^2(\Omega)}^2, \quad (5.7)$$

the minimizer f_k is a control such that the corresponding solution u_k satisfies $u_k(T) \rightarrow u^1$ as k tends to ∞ in $L^2(\Omega)$. In [69] a logarithmic convergence rate was proved for this procedure.

REMARK 5.4. As we mentioned above, the heat equation, despite the infinite speed of propagation behaves quite differently in unbounded domains. In [126] it was proved that, even if approximate controllability holds, null controllability does not hold for the heat equation in the whole line when the control acts in a bounded subdomain. But, approximate controllability holds, and can be even extended to semilinear equations [46,48,57]. Null controllability may be achieved when the support of the control is such that its complement is a bounded set. In that case the situation is fairly similar to the case where the equation holds in a bounded domain [22]. Similar results hold also in some particular cases in which the uncontrolled domain is unbounded but the distance to the controlled region is uniformly bounded (see [23] where the case of controls supported in a sequence of annulae is considered). We also refer to [133] for other results in this context. In particular, the possibility of getting the null controllability in an infinite cylinder is proved. The proof of the latter is based on an Fourier decomposition, allowing to reduce the problem to a family of heat equations of lower dimension that can be proved to be uniformly controllable by the methods presented above. This strategy was used in [174] to get uniform controllability

results for heat equations in thin domains and in [181] for proving convergence for finite-difference approximations in the context of the null controllability of the multidimensional heat equation. The paper [133] also contains an interesting example of an infinite rod which fails to be controllable, despite of the fact that the distance of all its points to the control region is uniformly bounded.

REMARK 5.5. In this section we have discussed heat equations with zeroth-order bounded potentials. Similar estimates, with different exponents, can be obtained when the potential belongs to L^p , for p large enough (see [53]). But we could also consider equations with convective terms. For instance,

$$\begin{cases} -\varphi_t - \Delta\varphi + \operatorname{div}(W(x, t)\varphi) + a(x, t)\varphi = 0 & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x) & \text{in } \Omega, \end{cases} \quad (5.8)$$

where $W = W(x, t)$ is a bounded convective potential (see [98]). The observability inequalities are true in this more general case, the observability constant being affected by an added term of the form $\exp(C\|W\|_\infty^2)$. This allows addressing control problems for semi-linear heat equations depending on the gradient. In this case the growth of the nonlinearity on ∇u has to be asymptotically smaller than $s \log^{1/2}(s)$ (see [4,50]).

5.2. Optimality

The observability constant in (5.1) includes three different terms. More precisely,

$$\begin{aligned} & \exp\left(C\left(1 + \frac{1}{T} + T\|a\|_\infty + \|a\|_\infty^{2/3}\right)\right) \\ &= C_1^*(T, a)C_2^*(T, a)C_3^*(T, a), \end{aligned} \quad (5.9)$$

where

$$\begin{aligned} C_1^*(T, a) &= \exp\left(C\left(1 + \frac{1}{T}\right)\right), \\ C_2^*(T, a) &= \exp(CT\|a\|_\infty), \\ C_3^*(T, a) &= \exp(C\|a\|_\infty^{2/3}). \end{aligned} \quad (5.10)$$

The role that each of these constants plays in the observability inequality is of different nature. It is roughly as follows:

- When $a \equiv 0$, i.e., in the absence of potential, the observability constant is simply $C_1^*(T, a)$. This constant blows up exponentially as $T \downarrow 0$. This growth rate is easily seen to be optimal by inspection of the heat kernel and has been analyzed in more detail in [69] and [131], in terms of the geometry of Ω and ω . We refer also to [152] for a discussion of the optimal growth rate for boundary observability in one space dimension.

- The second constant $C_2^*(T, a)$ is very natural as it arises when applying Gronwall's inequality to analyze the time evolution of the L^2 -norm of solutions. More precisely, it arises when getting (5.1) out of (5.2).
- The constant $C_3^*(T, a)$, which, actually, only depends on the potential a , is the most intriguing one. Indeed, the $2/3$ exponent does not seem to arise naturally in the context of the heat equation since, taking into account that the heat operator is of order one and two in the time and space variables respectively, one could rather expect terms of the form $\exp(c\|a\|_\infty)$ and $\exp(c\|a\|_\infty^{1/2})$, as a simple ODE argument would indicate.

In the recent paper [53] we show that, surprisingly, to some extent, the last contribution $C_3^*(T, a)$ to the observability constant is optimal. This happens for systems of two heat equations, in even space dimension and in the range of values of time T ,

$$\|a\|_\infty^{-2/3} \lesssim T \lesssim \|a\|_\infty^{-1/3}. \quad (5.11)$$

Here and in the remainder of this section we refer to systems of heat equations of the form

$$\begin{cases} -\varphi_t - \Delta\varphi + A(x, t)\varphi = 0 & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x) & \text{in } \Omega, \end{cases} \quad (5.12)$$

where $\varphi = \varphi(x, t)$ is a vector-valued function with N components and the potential $A = A(x, t)$ is matrix valued, with bounded measurable coefficients. System (5.12) is weakly coupled since it is diagonal in the principal part and it is only coupled through the zeroth-order term.

The Carleman inequality (5.4) yields for these systems observability inequalities of the form (5.1), following exactly the same method as for scalar equations. This is so because, as we said, the system is uncoupled in the principal part. In fact, as mentioned above, the Carleman inequality refers to the principal part of the operator, which in the present case is the heat operator componentwise, and the lower-order term is simply treated as a perturbation.

There is still a lot to be understood for these problems: scalar equations, one space dimension, other time intervals, etc. But the interest of this first optimality result is to confirm the need of the unexpected term $C_3^*(T, a)$ in the observability inequality (5.1). This is also relevant in view of applications to nonlinear problems, since it is this constant that determines the maximal growth of the nonlinearity for which null-controllability is known to hold in an uniform time: $s \log^{3/2}(s)$, [70].

The following holds.

THEOREM 5.2 [53]. *Assume that the space dimension $n \geq 2$ is even and that the number of equations of the parabolic system is $N \geq 2$. Let Ω be a bounded domain of $\mathbb{R}^n$ and ω a nonempty open subset of Ω . Then there exists $c > 0$, $\mu > 0$, a family $(A_R)_{R>0}$ of matrix-valued potentials such that*

$$\|A_R\| \xrightarrow{R \rightarrow +\infty} +\infty,$$

and a family $(\varphi_R^0)_{R>0}$ of initial conditions in $(L^2(\Omega))^N$ so that the corresponding solutions φ_R of (5.12) with $A = A_R$ satisfy

$$\lim_{R \rightarrow \infty} \left\{ \inf_{T \in I_\mu} \frac{\|\varphi_R(0)\|_{(L^2(\Omega))^N}^2}{\exp(c\|A_R\|_\infty^{2/3}) \int_0^T \int_\omega |\varphi_R|^2 dx dt} \right\} = +\infty, \quad (5.13)$$

where $I_\mu = (0, \mu\|A_R\|^{-1/3}]$.

Let us briefly sketch its proof. We refer to [53] for more details and other results and open problems related with this issue. In particular, in [53] the wave equation with lower-order terms is also considered and sharp observability inequalities are proved.

Theorem 5.2 is a consequence of the following known result.

THEOREM 5.3 (Meshkov [124]). *Assume that the space dimension is $n = 2$. Then, there exists a nonzero complex-valued bounded potential $q = q(x)$ and a nontrivial complex valued solution $u = u(x)$ of*

$$\Delta u = q(x)u \quad \text{in } \mathbb{R}^2, \quad (5.14)$$

with the property that

$$|u(x)| \leq C \exp(-|x|^{4/3}) \quad \forall x \in \mathbb{R}^2 \quad (5.15)$$

for some positive constant $C > 0$.

This construction by Meshkov provides a complex-valued bounded potential $q = q(x)$ in $\mathbb{R}^2$ and a nontrivial solution u of the elliptic equation (5.14) with the decay property $|u(x)| \leq \exp(-|x|^{4/3})$. This decay estimate turns out to be sharp as proved by Meshkov by Carleman inequalities. In other words, if, given a bounded potential q , the solution of (5.14) decays faster than $\exp(-C|x|^{4/3})$ for all $C > 0$ then, necessarily, this solution is the trivial one. Meshkov's construction may be generalized to any even dimension by separation of variables. We refer to [53] for a similar construction in odd dimension with a slightly weaker decay rate (essentially the same exponential decay up to a multiplicative logarithmic factor).

Theorem 5.2 holds from the construction by Meshkov by scaling and localization arguments. To simplify the presentation we focus in the case of two space dimensions $n = 2$ and of systems with two components $N = 2$ in which case Meshkov's result can be applied in a more straightforward way.

Its proof is divided into several steps.

STEP 1 (Construction on $\mathbb{R}^n$). Consider the solution u and potential q given by Theorem 5.3. By setting

$$u_R(x) = u(Rx), \quad A_R(x) = R^2 q(Rx), \quad (5.16)$$

we obtain a one-parameter family of potentials $\{A_R\}_{R>0}$ and solutions $\{u_R\}_{R>0}$ satisfying

$$\Delta u_R = A_R(x)u_R \quad \text{in } \mathbb{R}^n \quad (5.17)$$

and

$$|u_R(x)| \leq C \exp(-R^{4/3}|x|^{4/3}) \quad \text{in } \mathbb{R}^n. \quad (5.18)$$

These functions may also be viewed as stationary solutions of the corresponding parabolic systems. Indeed, $\psi_R(t, x) = u_R(x)$, $x \in \mathbb{R}^n$, $t > 0$, satisfies

$$\psi_{R,t} - \Delta \psi_R + A_R \psi_R = 0, \quad x \in \mathbb{R}^n, t > 0, \quad (5.19)$$

and

$$|\psi_R(x, t)| \leq C \exp(-R^{4/3}|x|^{4/3}), \quad x \in \mathbb{R}^n, t > 0. \quad (5.20)$$

STEP 2 (Restriction to Ω). Let us now consider the case of a bounded domain Ω and ω to be a nonempty open subset Ω such that $\omega \neq \Omega$. Without loss of generality (by translation and scaling) we can assume that $B \subset \Omega \setminus \bar{\omega}$.

We can then view the functions $\{\psi_R\}_{R>0}$ above as a family of solutions of the Dirichlet problem in Ω with nonhomogeneous Dirichlet boundary conditions:

$$\begin{cases} \psi_{R,t} - \Delta \psi_R + A_R \psi_R = 0 & \text{in } Q, \\ \psi_R = \varepsilon_R & \text{on } \Sigma, \end{cases} \quad (5.21)$$

where $\varepsilon_R = \psi_R|_{\partial\Omega} = u_R|_{\partial\Omega}$.

Taking into account that both ω and $\partial\Omega$ are contained in the complement of B , we deduce that, for a suitable C ,

$$|\psi_R(t, x)| \leq C \exp(-R^{4/3}), \quad x \in \omega, 0 < t < T, \quad (5.22)$$

$$|\varepsilon_R(t, x)| \leq C \exp(-R^{4/3}), \quad x \in \partial\Omega, 0 < t < T, \quad (5.23)$$

$$\|\psi_R(T)\|_{L^2(\Omega)}^2 \sim \|\psi_R(T)\|_{L^2(\mathbb{R}^n)}^2 = \|u_R\|_{L^2(\mathbb{R}^n)}^2 = \frac{1}{R^n} \|u\|_{L^2(\mathbb{R}^n)}^2 = \frac{c}{R^n}, \quad (5.24)$$

$$\|A_R\|_{L^\infty(\Omega)} \sim \|A_R\|_{L^\infty(\mathbb{R}^n)} = CR^2. \quad (5.25)$$

We can then correct these solutions to fulfill the Dirichlet homogeneous boundary condition. For this purpose, we introduce the correcting terms

$$\begin{cases} \rho_{R,t} - \Delta \rho_R + A_R \rho_R = 0 & \text{in } Q, \\ \rho_R = \varepsilon_R & \text{on } \Sigma, \\ \rho_R(0, x) = 0 & \text{in } \Omega, \end{cases} \quad (5.26)$$

and then set

$$\varphi_R = \psi_R - \rho_R. \quad (5.27)$$

Clearly $\{\varphi_R\}_{R>0}$ is a family of solutions of parabolic systems of the form (5.12) with potentials $A_R(x) = R^2 q(Rx)$.

The exponential smallness of the Dirichlet data ε_R shows that ρ_R is exponentially small too. This allows showing that φ_R satisfies essentially the same properties as ψ_R in (5.22)–(5.25). Thus, the family φ_R suffices to show that the statement in Theorem 5.2 holds.

6. Parabolic equations with low regularity coefficients

In this section we briefly discuss the problem of controllability for parabolic equations with low regularity coefficients in the principal part.

The same issue is relevant for wave equations too. In that case, according to the results in [26], we know that observability inequalities and exact controllability properties may fail for wave equations with Hölder continuous coefficients even in one space dimension. In 1- d we also know that exact controllability holds with BV -coefficients [40,71]. The picture is not complete in the multidimensional case, in which the various existing methods require different regularity properties. The method of multipliers requires coefficients to be C^1 or Lipschitz continuous because, when integrating by parts, one is forced to take one derivative of the coefficients in the principal part of the operator. Roughly speaking, the same happens for the Carleman inequality approach (although the Lipschitz condition can be replaced by a suitable L^p estimate on the first-order derivatives). Obviously, in both cases, other structural assumptions are needed on the coefficients (not only regularity) to guarantee that the observability inequality holds. The microlocal approach requires more regular coefficients. Indeed, $C^{1,1}$ coefficients are needed in order to prove existence, uniqueness and stability of bicharacteristic rays and, as far as we know, this is the only context in which the GCC is known to suffice. In fact, the extension of the GCC for less regular coefficients has not been formulated since, as we said, when coefficients fail to be $C^{1,1}$, the Hamiltonian system determining the bicharacteristic rays is not necessarily well posed.

Much less is known for parabolic operators. The Carleman inequality approach works for Lipschitz continuous coefficients. But we do not know whether this assumption is needed or not. Indeed, there is no counterexample in the literature justifying the need of regularity assumptions on the coefficients other than being merely bounded and measurable. A first result for piecewise constant coefficients by means of Carleman inequalities has been established in [51] but imposing monotonicity conditions on the interfaces. Indeed Carleman inequalities, as multipliers for wave equations, generate spurious terms on interfaces and, so far, the only way of getting rid of them is precisely imposing these monotonicity conditions on the interfaces to guarantee they have the good sign. For wave equations these conditions are known to be natural since they avoid trapped waves [123]. But in the context of heat equations there is no evidence of the need of such conditions. In [71] it

has been proved that BV -regularity of coefficients suffices in $1-d$. Moreover, in [1] it has been shown that the same is true for bounded measurable coefficients. This shows that the monotonicity conditions are not always required. But the methods in [1] and [71] only apply to coefficients depending only on x . Consequently, the problem of getting observability inequalities for heat equations with nonsmooth coefficients is widely open.

In this section we pursue a classical argument in the theory of PDE that consists in considering small L^∞ perturbations of a constant coefficient heat operator. The basic ingredient for doing that is the Carleman inequality (5.4) that not only yields estimates on the solution of the heat equation but also on the leading order terms. Before considering heat equations, in order to illustrate the methods, we discuss elliptic equations by means of the sharp Carleman inequalities proved in [96–98] and, more precisely, the problem of unique continuation of eigenfunctions.

6.1. Elliptic equations

We consider the elliptic problem

$$\begin{cases} -\Delta y = f + \sum_{j=1}^n \partial_j f_j & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega. \end{cases} \quad (6.1)$$

Let ω be an open nonempty subset of Ω and consider the weight function η^0 in (5.3). Set $\tilde{\rho}(x) = \exp(\exp(\lambda\eta^0(x)))$, where η^0 is as in (5.3). The following Carleman estimate was proved in [98] (see also [96,97] for an extension to nonhomogeneous boundary conditions).

THEOREM 6.1. *There exist positive constants $\widehat{C} > 0$, s_0 and λ_0 , which only depend on Ω and ω , such that for all $s \geq s_0$ and $\lambda \geq \lambda_0$ the following inequality holds for every solution of (6.1):*

$$\begin{aligned} & \int_{\Omega} [\tilde{\rho}^{2s} |\nabla y|^2 + s^2 \lambda^2 \exp(2\lambda\eta^0) \tilde{\rho}^{2s} |y|^2] dx \\ & \leq \widehat{C} \left[\frac{1}{s\lambda^2} \int_{\Omega} \frac{\tilde{\rho}^{2s}}{\exp(\lambda\eta^0)} f^2 dx \right. \\ & \quad + s \int_{\Omega} \exp(\lambda\eta^0) \tilde{\rho}^{2s} \sum_{j=1}^n |f_j|^2 dx \\ & \quad \left. + \int_{\omega} \tilde{\rho}^{2s} (|\nabla y|^2 + s^2 \lambda^2 \exp(2\lambda\eta^0) |y|^2) dx \right]. \end{aligned} \quad (6.2)$$

In [96] and extension of this result has been proved including elliptic problems with variable smooth coefficients in the principal part and nonhomogeneous Dirichlet data in $H^{1/2}(\partial\Omega)$. Here, for the sake of simplicity, we restrict our attention to the case of homogeneous boundary conditions. Strictly speaking, by viewing the solutions of (6.1) as

time-independent solutions of the corresponding parabolic problem, in the present case, inequality (6.2) is a consequence of the parabolic inequalities in [98].

Note that the estimate is sharp in what concerns the order of the different terms entering in it. Indeed, an estimate of the right-hand side term in $H^{-1}(\Omega)$ allows getting estimates on the solutions in $H^1(\Omega)$ in appropriate weighted norms. In fact, the proof of Theorem 6.1 requires of Carleman estimates and duality arguments to deal with right-hand side terms in H^{-1} .

Let us now consider an elliptic operator with bounded coefficients. To simplify the presentation we consider the case

$$\begin{cases} -\operatorname{div}((1 + \varepsilon(x))\nabla y) = f + \sum_{j=1}^n \partial_j f_j & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega, \end{cases} \quad (6.3)$$

where ε is assumed to belong to $L^\infty(\Omega)$ and small so that $\|\varepsilon\|_{L^\infty(\Omega)} < 1$. This guarantees the ellipticity of the operator.

In order to extend the Carleman inequality to this class of operators, it is natural to view the leading term as follows:

$$-\operatorname{div}((1 + \varepsilon(x))\nabla y) = -\Delta y - \operatorname{div}(\varepsilon(x)\nabla y).$$

We can then rewrite (6.3) in the form (6.1) but with f_j replaced by $\tilde{f}_j = f_j + \varepsilon(x) \partial_j y$. When doing that and applying (6.2), we deduce that

$$\begin{aligned} & \int_{\Omega} [\tilde{\rho}^{2s} |\nabla y|^2 + s^2 \lambda^2 \exp(2\lambda\eta^0) \tilde{\rho}^{2s} |y|^2] \, dx \\ & \leq \widehat{C} \left[\frac{1}{s\lambda^2} \int_{\Omega} \frac{\tilde{\rho}^{2s}}{\exp(\lambda\eta^0)} f^2 \, dx \right. \\ & \quad + s \int_{\Omega} \exp(\lambda\eta^0) \tilde{\rho}^{2s} \sum_{j=1}^n (|f_j|^2 + \|\varepsilon\|_{L^\infty(\Omega)}^2 |\partial_j y|^2) \, dx \\ & \quad \left. + \int_{\omega} \tilde{\rho}^{2s} (|\nabla y|^2 + s^2 \lambda^2 \exp(2\lambda\eta^0) |y|^2) \, dx \right]. \end{aligned} \quad (6.4)$$

Thus, with respect to (6.2) this adds the extra term

$$\widehat{C} s \int_{\Omega} \exp(\lambda\eta^0) \tilde{\rho}^{2s} \|\varepsilon\|_{L^\infty(\Omega)}^2 |\nabla y|^2 \, dx.$$

But this term can be absorbed by the left-hand side term in (6.4) provided

$$\widehat{C} s \exp(\lambda\|\eta^0\|_{L^\infty(\Omega)}) \|\varepsilon\|_{L^\infty(\Omega)}^2 < 1. \quad (6.5)$$

To be more precise, assume that ε satisfies the smallness condition

$$\|\varepsilon\|_{L^\infty(\Omega)}^2 < \delta \widehat{C} s_0^{-1} \exp\left(-\lambda_0 \max_{\Omega}(\eta^0)\right) \quad (6.6)$$

for some $\delta < 1$. Then (6.2) holds also for the solutions of (6.3), with a larger constant $C = \widehat{C}/\delta > 0$, where $\widehat{C}$ is the one in (6.2).

Accordingly, the following theorem holds.

THEOREM 6.2. *Assume that ε satisfies the smallness condition (6.6) with $\delta < 1$ and where $\widehat{C} > 0$, s_0 and λ_0 are as in (6.2). Then (6.2) holds for the solutions of (6.3) for a larger observability constant $\widehat{C}/\delta > 0$.*

Whether the smallness condition (6.6) is needed for (6.2) to hold or not is an open problem.

Note that, in particular, these Carleman inequalities may be used to prove unique continuation properties. Let us consider for instance the spectral problem:

$$\begin{cases} -\operatorname{div}((1 + \varepsilon(x))\nabla y) = \gamma^2 y & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega. \end{cases} \quad (6.7)$$

When applying the inequality to the solution of (6.7) we get

$$\begin{aligned} & \int_{\Omega} [\tilde{\rho}^{2s} |\nabla y|^2 + s^2 \lambda^2 \exp(2\lambda\eta^0) \tilde{\rho}^{2s} |y|^2] \, dx \\ & \leq C \left[\frac{\gamma^4}{s\lambda^2} \int_{\Omega} \frac{\tilde{\rho}^{2s}}{\exp(\lambda\eta^0)} y^2 \, dx \right. \\ & \quad + s \int_{\Omega} \exp(\lambda\eta^0) \tilde{\rho}^{2s} \sum_{j=1}^n (\|\varepsilon\|_{L^\infty(\Omega)}^2 |\partial_j y|^2) \, dx \\ & \quad \left. + \int_{\omega} \tilde{\rho}^{2s} (|\nabla y|^2 + s^2 \lambda^2 \exp(2\lambda\eta^0) |y|^2) \, dx \right]. \end{aligned} \quad (6.8)$$

Absorbing the two remainder terms

$$\widehat{C} \frac{\gamma^4}{s\lambda^2} \int_{\Omega} \frac{\tilde{\rho}^{2s}}{\exp(\lambda\eta^0)} y^2 \, dx$$

and

$$s\widehat{C} \int_{\Omega} \exp(\lambda\eta^0) \tilde{\rho}^{2s} \sum_{j=1}^n (\|\varepsilon\|_{L^\infty(\Omega)}^2 |\partial_j y|^2) \, dx$$

on the right-hand side requires the smallness condition (6.6) together with the bound

$$\widehat{C}\gamma^4 < s^3\lambda^4 \exp\left(3\lambda \min_{\Omega} \eta^0\right) = s^3\lambda^4. \quad (6.9)$$

This means that in order to cover larger and larger eigenfrequencies γ one is required to assume that the perturbation ε is smaller and smaller.

As we shall see this limitation appears also when dealing with evolution problems. This is natural indeed, since, when considering evolution problems all the spectrum of the underlying elliptic operator is involved simultaneously.

Once the smallness conditions (6.6) and (6.9) are imposed one guarantees the following observability inequality to hold:

$$\begin{aligned} & \int_{\Omega} \tilde{\rho}^{2s} [|\nabla y|^2 + s^2\lambda^2 \exp(2\lambda\eta^0) \tilde{\rho}^{2s} |y|^2] dx \\ & \leq C \int_{\omega} \tilde{\rho}^{2s} (|\nabla y|^2 + s^2\lambda^2 \exp(2\lambda\eta^0) |y|^2) dx, \end{aligned} \quad (6.10)$$

for some $C > 0$. This implies, in particular, the property of unique continuation: *If $y \equiv 0$ in ω , then, necessarily, $y \equiv 0$ everywhere.*

At this respect it is important to note that, in what concerns unique continuation, in two space dimensions, this property is guaranteed for bounded measurable coefficients without further regularity assumptions ([10] and [11]). But the techniques of proof are specific to 2- d . However these results do not provide quantitative estimates as those we obtained above.

The situation is totally different in higher dimensions. Indeed, for $n \geq 3$ it is well known that unique continuation fails, in general, for elliptic equations with measurable (and even Hölder continuous) coefficients (see [129] and, for elliptic equations in nondivergence form, [141]). Therefore it is natural that the methods we develop here, based on global Carleman inequalities, that do not distinguish the various space dimensions, require restrictions on the size of the bounded measurable perturbations of the coefficients allowed.

The situation is different for equations in which the density is perturbed. In this case the corresponding eigenvalue problem reads

$$\begin{cases} -\Delta y = \gamma^2(1 + \varepsilon(x))y & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega. \end{cases} \quad (6.11)$$

In this case one gets (6.10) under a suitable smallness assumption on ε , but, contrarily to the elliptic problem (6.7), for all the spectrum simultaneously.

In the following subsection we apply the same ideas to a parabolic equation with bounded small perturbations in the leading density coefficient.

6.2. Parabolic equations

Let us now consider the following parabolic equation

$$\begin{cases} -(1 + \varepsilon(x, t))\varphi_t - \Delta\varphi = 0 & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x) & \text{in } \Omega. \end{cases} \quad (6.12)$$

This is an adjoint heat equation with a variable density $\rho(x, t) = 1 + \varepsilon(x, t)$. We assume that $\varepsilon \in L^\infty(\Omega \times (0, T))$ satisfies the smallness condition $\|\varepsilon\|_{L^\infty(\Omega \times (0, T))} < 1$ so that the system is parabolic and well posed.

As in the elliptic case it is natural to decompose the parabolic operator into the heat one plus a small perturbation. The heat equation in (6.12) can then be written in the form

$$[-\varphi_t - \Delta\varphi] = \varepsilon(x, t)\varphi_t.$$

Applying the Carleman inequality (5.4) to the solution of (6.12) we get

$$\begin{aligned} & \frac{1}{s} \int_Q \rho^{-2s} t(T-t) [|\varphi_t|^2 + |\Delta\varphi|^2] dx dt \\ & + s \int_Q \rho^{-2s} t^{-1} (T-t)^{-1} |\nabla\varphi|^2 dx dt + s^3 \int_Q \rho^{-2s} t^{-3} (T-t)^{-3} \varphi^2 dx dt \\ & \leq C_* \left[\int_Q \rho^{-2s} |\varepsilon(x, t)\varphi_t|^2 dx dt \right. \\ & \quad \left. + s^3 \int_0^T \int_\omega \rho^{-2s} t^{-3} (T-t)^{-3} \varphi^2 dx dt \right]. \end{aligned} \quad (6.13)$$

The term $\int_Q \rho^{-2s} |\varepsilon(x, t)\varphi_t|^2 dx dt$ on the right-hand side can be viewed as a remainder. In order to get rid of it and to get an observability estimate for the solutions of the perturbed system (6.12) we need to assume that

$$\frac{\varepsilon^2(x, t)}{t(T-t)} \leq \frac{s}{C_*}. \quad (6.14)$$

This is clearly a smallness condition on the perturbation ε of the constant coefficient. But it also imposes that ε vanishes at $t = 0$ and $t = T$. Indeed, in order to see this it is better to write

$$\varepsilon(x, t) = \sqrt{t}\sqrt{T-t}\delta(x, t). \quad (6.15)$$

Then, the smallness condition reads

$$\|\delta\|_{L^\infty(Q)}^2 \leq \frac{s}{C_*}. \quad (6.16)$$

According to this analysis, the following holds.

THEOREM 6.3. *Let Ω be a bounded smooth domain and ω a nonempty open subset. Let $T > 0$. Let also s be large enough so that (5.4) holds with constant C_* . Consider the variable density heat equation (6.12) with ε as in (6.15) and satisfying the smallness condition (6.16). Then, the observability inequalities (5.1) and (5.2) hold for the solutions of (6.12).*

Strictly speaking, the arguments above provide a Carleman inequality of the form (5.4) for the solutions of (6.12). Estimates of the form (5.1) and (5.2) can then be obtained following classical arguments (see [69]). One first derives (5.2) as an immediate consequence of the Carleman inequality to later obtain (5.1) as a consequence of the well-posedness of (6.12). Indeed, multiplying in (6.12) by φ_t and integrating by parts we deduce the energy identity

$$\frac{1}{2} \frac{d}{dt} \int_{\Omega} |\nabla \varphi|^2 dx = \int_{\Omega} (1 + \varepsilon(x, t)) |\varphi_t|^2 dx \geq 0.$$

Thus

$$\int_{\Omega} |\nabla \varphi|^2(x, t) dx \leq \int_{\Omega} |\nabla \varphi|^2(x, T) dx$$

for all $0 \leq t \leq T$. In view of this and, as a consequence of the Carleman inequality (5.4), we deduce that

$$\int_0^T \int_{\Omega} e^{-A/(T-t)} |\nabla \varphi|^2 dx dt \leq C \int_0^T \int_{\omega} \varphi^2 dx dt \quad (6.17)$$

and

$$\int_{\Omega} |\nabla \varphi(x, 0)|^2 dx dt \leq C \int_0^T \int_{\omega} \varphi^2 dx dt \quad (6.18)$$

for suitable constants $C, A > 0$. By Poincaré inequality, this yields

$$\int_0^T \int_{\Omega} e^{-A/(T-t)} \varphi^2 dx dt \leq C \int_0^T \int_{\omega} \varphi^2 dx dt \quad (6.19)$$

and

$$\int_{\Omega} \varphi^2(x, 0) dx dt \leq C \int_0^T \int_{\omega} \varphi^2 dx dt. \quad (6.20)$$

Whether the smallness conditions (6.15) and (6.16) on ε are needed or not is an open problem. It is however convenient to distinguish between $t = 0$ and $t = T$.

The restriction at $t = 0$ can be relaxed. Indeed, using the regularizing effect of the heat equation, a variant of (5.4) can be obtained so that the weight function involved in it does not degenerate at $t = 0$ (see [76]). In this way, a similar result would hold under the restriction

$$\frac{\varepsilon^2(x, t)}{T - t} \leq \frac{s}{C_*}, \quad (6.21)$$

instead of (6.14).

At the contrary getting rid of the condition that $\varepsilon^2(x, t)/(T - t)$ is small at $t = T$ is probably very difficult. This smallness condition could even be necessary. This is related to the nature of the observability inequalities for the adjoint heat equation that are unable to provide estimates on the solutions at $t = T$ because of the very strong smoothing effect. In fact, as indicated in the context of the constant-coefficient heat equation, even when estimating the L^2 -norm of the solutions, an exponentially vanishing weight is needed at $t = T$ (see (5.2)).

This difficulty is probably also related to the one we encountered in the previous subsection when dealing with the spectrum of the system. There, we could only deal with an eigenvalue range whose width depended on the smallness condition on the perturbation of the coefficient. As indicated there, the perturbation needed to be smaller and smaller to be able to cover eventually the whole range of frequencies. Obviously, when dealing with the evolution problem the whole range of frequencies is involved. It is therefore natural to require an smallness assumption that vanishes as $t \rightarrow T$.

The observability result for the adjoint heat equation (6.12) we have proved yields immediately results on the null-controllability of the corresponding state equation

$$\begin{cases} \partial_t((1 + \varepsilon(x, t))u) - \Delta u = f \mathbb{1}_\omega & \text{in } Q, \\ u = 0 & \text{on } \Sigma, \\ u(x, 0) = u^0(x) & \text{in } \Omega. \end{cases} \quad (6.22)$$

The following holds.

THEOREM 6.4. *Under the assumptions of Theorem 6.3 system (6.22) is null-controllable.*

The control for (6.22) can be obtained from the variational methods we have developed in the context of the constant-coefficient heat equation. It may be built being of minimal L^2 -norm or of minimal L^∞ -norm, in which case it will be of quasibang-bang form.

As in Theorem 6.3, Theorem 6.4 requires smallness conditions of the form (6.15), (6.16) on the coefficient ε . As we said above, getting rid of this smallness condition at $t = T$ is probably a very difficult problem. This can also be interpreted in the context of the control of the state equation (6.22). Indeed, since we are trying to drive the state u to rest at time $t = T$, the oscillations of the density coefficients ε could be a major obstacle for doing that. Note however that, as mentioned in the introduction of this section, there is no example of heat equation with bounded coefficients for which the null controllability property fails. The situation is different in what concerns the degeneracy condition of the coefficient at

$t = 0$. As we said above, this condition is probably unnecessary for observability and for null controllability too, but this remains to be investigated.

A similar analysis could be developed for heat equations with bounded small perturbations on the coefficients entering in the second-order operator, i.e., for equations of the form

$$\begin{cases} -(1 + \varepsilon(x, t))\varphi_t - \operatorname{div}((1 + \sigma(x, t))\nabla\varphi) = 0 & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x) & \text{in } \Omega, \end{cases} \quad (6.23)$$

and

$$\begin{cases} \partial_t((1 + \varepsilon(x, t))u) - \operatorname{div}((1 + \sigma(x, t))\nabla u) = f\mathbb{1}_\omega & \text{in } Q, \\ u = 0 & \text{on } \Sigma, \\ u(x, 0) = u^0(x) & \text{in } \Omega. \end{cases} \quad (6.24)$$

For that purpose we need a parabolic version of the elliptic inequalities in the previous section. This was developed in [98] (see also [64]). Let us recall it briefly. Consider the adjoint heat equation

$$\begin{cases} -\varphi_t - \Delta\varphi = f + \sum_{j=1}^n \partial_j f_j & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = \varphi^0(x) & \text{in } \Omega. \end{cases} \quad (6.25)$$

The following holds [64,98]:

$$\begin{aligned} & \int_Q \rho^{-2s} \left[s\lambda^2 \frac{\exp(\lambda\eta^0)}{(t(T-t))^3} |\nabla\varphi|^2 + s^3\lambda^4 \frac{\exp(3\lambda\eta^0)}{(t(T-t))^3} |\varphi|^2 \right] dx dt \\ & \leq C \left[\int_Q \rho^{-2s} f^2 dx dt + s^2\lambda^2 \int_Q \frac{\exp(2\lambda\eta^0)}{(t(T-t))^2} \rho^{-2s} \sum_{j=1}^n |f_j|^2 dx dt \right. \\ & \quad \left. + s^3\lambda^4 \int_{\omega \times (0, T)} \rho^{-2s} \frac{\exp(3\lambda\eta^0)}{(t(T-t))^3} |\varphi|^2 dx dt \right]. \end{aligned} \quad (6.26)$$

As in the elliptic case, this inequality allows absorbing the effect of the term that the bounded perturbation on the principal part of the operator in (6.25) adds. More precisely, the term $\operatorname{div}(\sigma(x, t)\nabla\varphi)$ can be absorbed by a suitable smallness condition on σ ,

$$Cs \max_Q \exp(\lambda\eta^0) \frac{|\sigma(x, t)|^2}{t(T-t)} < 1, \quad (6.27)$$

which is similar to (6.14). Once more the smallness condition may be relaxed so that the vanishing weight is not needed at $t = 0$.

In fact these parabolic estimates contain those we have obtained in the previous subsection on elliptic equations. Indeed, by viewing solutions of the elliptic equations as time

independent solutions of the parabolic ones it is easy to see that the elliptic estimates are contained in the parabolic ones under similar smallness conditions on the perturbations of the coefficients.

In order to deal with the general parabolic equation (6.25) and to address variable densities ε and coefficients σ in the second-order elliptic operator we have to combine (5.4) and (6.26). In this way we get the Carleman inequality under similar smallness conditions both on ε and σ .

As an immediate consequence of these results we deduce the null controllability of the state equation (6.24).

REMARK 6.1. In [66] similar developments have been done in the context of the adjoint equation

$$-\partial_t((1 + \varepsilon(x, t))\varphi) - \Delta\varphi = 0.$$

Note that, in this case, the perturbation has to be viewed as an element of $H^{-1}(0, T; L^2(\Omega))$ what adds further technical difficulties. Some applications to the controllability of quasilinear parabolic problems have also been given. These two issues are closely related, as in the semilinear case, because of the use of the fixed point method which reduces the control of the nonlinear problem to a sharp estimate of the cost of controllability of the linearized one.

Control of quasilinear heat equations is also a widely open subject of research. Very likely, the approach based on linearization and a sharp analysis of the cost of controlling linear systems is insufficient to cover the new phenomena that quasilinear equations present.

REMARK 6.2. It is important to observe that the approach we have developed in this section requires smallness conditions on the perturbations of the coefficients. In particular, we could consider coefficients that, for all t are piecewise constant and possibly discontinuous. The results in this section do not require the monotonicity conditions in [51]. But, at the contrary, they need the jumps to be small, and to vanish as t tend to T at order $\sqrt{T-t}$.

It is also important to observe that the results of this section apply to coefficients which depend both on space and time, a framework that might be much richer than that of parabolic equations with low regularity coefficients depending only on x . For instance, as indicated above, in [1] the 1- d heat equation with bounded measurable coefficients depending only on x is null controllable. But it is unknown whether the same result is true for bounded measurable coefficients depending both on x and t .

7. Fluid–structure interaction models

7.1. Problem formulation

So far we have only discussed two model systems: the heat and the wave equation. But most of the techniques we have developed are also useful to address more sophisticated

and realistic models. That is the case, for instance, in the context of the Navier–Stokes equations in which most of the developments we have presented based on Carleman and observability inequalities, duality and variational principles allow obtaining a number of controllability results. It is by now well known, for instance, that the Navier–Stokes equations are locally null controllable (see, for instance, [34,36,68,73,75], and the references therein). Despite of the fact that the techniques described above apply, important further developments are needed to deal with the pressure term, the incompressibility condition, the lack of regularity of the convective potentials when linearizing the system around weak solutions, etc. We refer to the survey article [67] for an updated discussion of these issues. Controllability also holds for the Euler equations in an appropriate functional and geometric setting. However, because of the lack of viscosity, Carleman inequalities do not apply and different techniques have to be applied. We refer for instance to [33] and to [80] where the problem is solved in $2-d$ and $3-d$, respectively, by the so called “return method” due to Coron [35].

Much less is known in the context of fluid–structure interaction models. These models are indeed very hard to deal with because of their mixed hyperbolic–parabolic nature. Roughly speaking they can be viewed as the coupling of a Navier–Stokes system for the fluid, with a system of elasticity for the structure, coupled along a moving interface determined by the boundary of the deformed elastic body. In fact, even the problem of the well-posedness of these problems is badly understood. We refer to [13,14] and [39] for some results in that direction. The inviscid case has been also treated (see [137,138]). To the best of our knowledge, nothing is known on the controllability of this system.

This model may be simplified by assuming that the structure is a rigid body. In that case the modeling consists in coupling the Navier–Stokes equations to the ordinary differential equations for the motion of the rigid body. In $1-d$ in which a fluid modeled by the Burgers equation is coupled with a finite number of mass points, existence and uniqueness of global solutions is known. In particular, it is known that two solid particles may not collide in finite time, a problem that is still open in several space dimensions [91] and [161]. Recently, the first relevant results have been obtained at this respect. In [90] and [91] it has been proved independently that in the half plane, a spherical rigid body can not reach the boundary in finite time. Some other geometries have also been considered but the problem is still open in full generality. The same can be said about the possible collision of two moving rigid bodies. For this model, in $1-d$, in the presence of one single particle and with controls on both sides, null controllability has been proved in [49]. The difficulty one encounters when dealing with this apparently simple system from a control theoretical point of view is similar to that we found when considering heat equations with piecewise constant and discontinuous coefficients: it is hard to derive observability estimates by means of Carleman inequalities because of the interface terms. In particular, the problem of control of this $1-d$ fluid-mass model is open when the control acts on one side of the mass only. Recently a very interesting result has been obtained in $2-d$ in [15]. It guarantees the local null controllability for the Navier–Stokes equations coupled with the motion of a finite number of rigid bodies, the control being applied on an arbitrary open subset of the fluid and in an arbitrarily small time. In this sense the result is better than in $1-d$ where the control is assumed to be on both sides of the mass.

There is however an intuitive explanation of this fact. In 2- d the fluid envelopes the rigid bodies so that the effect of local controls on the fluid may propagate to all of it and then act on the masses too, while in 1- d the control, when applied only on one side, needs to cross the point-mass. The problem of controlling large initial data is still open. It is related to the fact that, as mentioned above, in general, in 2- d it is not known whether two rigid bodies may collide or whether they may collide with the external boundary. The estimates that Carleman inequalities yield become singular when collision occurs. Analyzing whether the control may avoid collision to occur (in case it occurs in the absence of control) and controlling to zero large initial data is certainly a very interesting and difficult problem.

These models are free boundary ones. When linearizing them around the equilibrium configuration they become evolution equations on two adjacent domains separated by a fixed interface. In this section we summarize some recent results on the asymptotic behavior of a linearized model arising in fluid–structure interaction, where a wave and a heat equation evolve in two adjacent bounded domains, with natural transmission conditions at the interface. The content of this section is based on joint work with Rauch and Zhang [146,165–167,169].

The system under consideration may be viewed as an approximate and simplified model for the motion of an elastic body immersed in a fluid, which, as we mentioned above, in its most rigorous modeling should be a nonlinear free boundary problem, with a moving interface between the fluid and the elastic body.

In the model we consider here the heat unknown is coupled with the velocity of the wave solution along the interface. A slightly simpler case in which the states of the heat and wave equations are directly coupled was addressed in [146]. Note however that the coupling conditions we consider here are more natural from the point of view of fluid–structure interaction.

In this section we mainly discuss the problem of the decay of solutions as t tends to zero. A similar study has been undertaken previously for the system of thermoelasticity (see [109]), another natural situation in which wave and heat equations are coupled. Note however that, in thermoelasticity, both the heat and the wave equation evolve in the same domain, while in the fluid–structure interaction model under consideration they evolve on two different domains, separated through an interface.

The model we consider here can be viewed as the coupling of the purely conservative dynamics generated by the wave equation and the strongly dissipative one that the heat equation produces. The total energy of solutions, addition of the thermal and elastic one, is dissipated through the heat domain. Therefore, studying the rate of decay of solutions of the whole system, is a way of addressing the issue of how strongly the two dynamics are coupled. Indeed, one could expect that, in case the two components of the system are coupled strongly enough along the interface, then solutions should decay with an exponential rate. This corresponds to the situation in which the semigroup $S(\cdot)$ generated by the system, which is dissipative, is such that $\|S(T)\| < 1$ for some $T > 0$ in the norm of the energy space. At the contrary, the lack of uniform exponential decay could be considered as a proof of the lack of strong coupling.

But the issue is more complex. Indeed, it is well known that, when the damping introduced on a wave-like equation is too strong, overdamping phenomena may occur and the exponential decay may be lost. This is for instance the case for the damped wave equation

$$u_{tt} - \Delta u + ku_t = 0,$$

with homogeneous Dirichlet boundary conditions. In this case the energy of solutions is given by

$$E(t) = \frac{1}{2} \int [|u_t|^2 + |\nabla u|^2] dx,$$

and the energy dissipation law

$$\frac{d}{dt} E(t) = -k \int |u_t|^2 dx.$$

In view of this energy dissipation law one could expect a faster decay rate for larger values of the dissipation parameter $k > 0$. But the exponential decay rate is not monotonic on $k > 0$. Indeed, despite the exponential decay rate increases as $k > 0$ is increasing and small, the decay rate diminishes when $k \rightarrow \infty$.

The damping that the heat equation introduces is an unbounded perturbation of the wave dynamics. This predicts that the exponential decay may be lost. This is indeed the case and it is independent of the geometry of the subdomain in which the heat and wave equations hold. In the case where the domain Ω is a polygon and the interface is a hyperplane, the lack of exponential decay can be proved by means of a plane wave analysis that allows exhibiting a class of solutions whose energy is mainly concentrated in the wave domain and therefore, very weakly dissipated through the heat mechanism. In general domains and for curved interfaces this construction needs of a more careful development based on the use of Gaussian beams.

But, on the other hand, due to the presence of the wave motion, and in view of our experience on the control and stabilization of the wave equation, one expects that the system will be more stable when the heat domain satisfies the GCC and, more precisely, when all rays in the wave domain enter the heat one in a uniform time. This is indeed the case. When the heat subdomain satisfies this GCC the decay rate of smooth solutions is polynomial but, in general, one can only guarantee a polynomial decay rate.

The main conclusions of the series of works we have mentioned above are roughly as follows:

- Whenever the heat subdomain is nonempty, the energy of solutions tends to zero as time tends to infinity.
- The decay rate is never exponentially uniform, regardless of the geometric properties of the heat subdomain. In other words, the dissipative semigroup generated by the system is of unit norm for all $t > 0$.
- When the heat domain satisfies the GCC, then the energy of smooth solutions decays polynomially.

- When the heat domain does not satisfy the GCC smooth solutions decay logarithmically for the simplified interface conditions in which continuity of the states is imposed. The problem is open for the more natural boundary conditions we shall consider here.

Some other issues are by now also well understood. In 1- d the problem of controllability has been solved in [165,166,177]. When the control acts on the exterior boundary of the wave domain null controllability can be easily proved using sidewise energy estimates for the wave equation and Carleman inequalities for the heat one. However, when the control acts on the extreme of the heat domain the space of controllable data is very small. Roughly, the controllable initial data have exponentially small Fourier coefficients on the basis of the eigenfunctions of the generator of the semigroup. In fact in [165] and [166] a complete asymptotic analysis of the spectrum of the system has been developed. According to it, the spectrum can be decomposed in two branches: the parabolic one and the hyperbolic one. The parabolic eigenvalues are asymptotically real and tend to $-\infty$ and the energy of the corresponding eigenfunctions is more and more concentrated on the heat domain. The hyperbolic one has vanishing asymptotic real part and their energy is concentrated on the wave domain. As a consequence of this fact the high frequency hyperbolic eigenfunctions are very badly controlled from the heat domain. Thus, for controlling a given initial datum, its Fourier components on the hyperbolic eigenfunctions need to vanish exponentially at high frequencies. These results, which have been obtained by means of 1- d methods are completely open in several space dimensions. In fact, according to them one also expects important differences for the multidimensional problem. In this section we only address the problem of the rate of decay. We do it in the multidimensional case. Thus, we do not employ spectral methods. Rather we use plane wave analysis and Gaussian beams to show that, whatever the geometry is, the decay rate is never exponential, and the existing hyperbolic observability estimates to prove the polynomial decay under the GCC. Although the results we present here are far from answering to the problem of controllability, by combining our understanding of the stabilization problem, and the behavior of the control problem in 1- d , one may at least guess what kind of results should be expected for the control problem in multi- d . It is natural for instance to expect that, if the control enters in a subset of the wave domain satisfying the GCC, then one should expect null controllability on the energy space. However, when controlling on the heat subdomain the space of controllable data should be very small, even if the heat domain envelops the wave one and satisfies the GCC. This is a widely open subject of research. In one space dimension other closely related models have also been investigated. In particular, a model coupling the wave equation with an equation of viscoelasticity (see [120]).

7.2. The model

Let $\Omega \subset \mathbb{R}^n$ ($n \in \mathbb{N}^*$) be a bounded domain with C^2 boundary $\Gamma = \partial\Omega$. Let Ω_1 be a subdomain of Ω and set $\Omega_2 = \Omega \setminus \overline{\Omega}_1$. We denote by γ the interface, by $\Gamma_j = \partial\Omega_j \setminus \bar{\gamma}$, $j = 1, 2$, the exterior boundaries, and by ν_j the unit outward normal vector of Ω_j , $j = 1, 2$. We assume $\gamma \neq \emptyset$ and γ is of class C^1 .

Consider the following hyperbolic–parabolic coupled system:

$$\begin{cases} y_t - \Delta y = 0 & \text{in } (0, \infty) \times \Omega_1, \\ z_{tt} - \Delta z = 0 & \text{in } (0, \infty) \times \Omega_2, \\ y = 0 & \text{on } (0, \infty) \times \Gamma_1, \\ z = 0 & \text{on } (0, \infty) \times \Gamma_2, \\ y = z_t, \quad \frac{\partial y}{\partial \nu_1} = -\frac{\partial z}{\partial \nu_2} & \text{on } (0, \infty) \times \gamma, \\ y(0) = y_0 & \text{in } \Omega_1, \\ z(0) = z_0, \quad z_t(0) = z_1 & \text{in } \Omega_2. \end{cases} \quad (7.1)$$

As we said before, this is a simplified and linearized model for fluid–structure interaction. In system (7.1), y may be viewed as the velocity of the fluid; while z and z_t represent respectively the displacement and velocity of the structure. More realistic models should involve the Stokes (resp. the elasticity) equations instead of the heat (resp. the wave) ones.

In [146] and [167], the same system was considered but for the transmission condition $y = z$ on the interface instead of $y = z_t$. But, from the point of view of fluid–structure interaction, the transmission condition $y = z_t$ in (7.1) is more natural. Note also that, as indicated above, the interface in this model is fixed. This corresponds to the fact that the system is a linearization around the trivial solution of a free boundary problem.

Set $H_{\Gamma_1}^1(\Omega_1) \triangleq \{h|_{\Omega_1} \mid h \in H_0^1(\Omega)\}$ and $H_{\Gamma_2}^1(\Omega_2) \triangleq \{h|_{\Omega_2} \mid h \in H_0^1(\Omega)\}$. System (7.1) is well posed in the Hilbert space

$$H \triangleq L^2(\Omega_1) \times H_{\Gamma_2}^1(\Omega_2) \times L^2(\Omega_2).$$

When Γ_2 is a nonempty open subset of the boundary (or, more generally, of positive capacity), in H the following norm is equivalent to the canonical one:

$$\begin{aligned} \|f\|_H &= \left[\|f_1\|_{L^2(\Omega_1)}^2 + \|\nabla f_2\|_{(L^2(\Omega_2))^n}^2 + \|f_3\|_{L^2(\Omega_2)}^2 \right]^{1/2} \\ \forall f &= (f_1, f_2, f_3) \in H. \end{aligned}$$

In this case the only stationary solution is the trivial one. This is due to the fact that Poincaré inequality holds.

When Γ_2 vanishes, $\|\cdot\|_H$ is no longer a norm on H . In this case, there are nontrivial stationary solutions of the system. Thus, the asymptotic behavior is more complex and one should rather expect the convergence of each individual trajectory to a specific stationary solution. To simplify the presentation in this section we assume that the capacity of Γ_2 is positive.

The energy of system (7.1) is given by

$$E(t) \triangleq E(y, z, z_t)(t) = \frac{1}{2} \|(y(t), z(t), z_t(t))\|_H^2$$

and satisfies the dissipation law

$$\frac{d}{dt}E(t) = - \int_{\Omega_1} |\nabla y|^2 dx. \quad (7.2)$$

Therefore, the energy of (7.1) decreases as $t \rightarrow \infty$.

In fact $E(t) \rightarrow 0$ as $t \rightarrow \infty$, without any geometric conditions on the domains Ω_1 and Ω_2 (other than the capacity of Γ_2 being positive). However, due to the lack of compactness of the domain of the generator of the underlying semigroup of system (7.1) for $n \geq 2$, one cannot use directly LaSalle's invariance principle to prove this result. Instead, using the "relaxed invariance principle" [154], we conclude that y and z_t tend to zero strongly in $L^2(\Omega_1)$ and $L^2(\Omega_2)$, respectively; while z tends to zero weakly in $H_{\Gamma_2}^1(\Omega_1)$ as $t \rightarrow \infty$. Then, we use the special structure of (7.1) and the key energy dissipation law (7.2) to obtain the strong convergence of z in $H_{\Gamma_2}^1(\Omega_1)$ [169].

Once the energy of each individual trajectory has been shown to tend to zero as t goes to ∞ , we analyze the rate of decay. In particular, it is natural to analyze whether there is an uniform exponential decay rate, i.e., whether there exist two positive constants C and α such that

$$E(t) \leq CE(0)e^{-\alpha t} \quad \forall t \geq 0, \quad (7.3)$$

for every solution of (7.1).

According to the energy dissipation law (7.2), the uniform decay problem (7.3) is equivalent to showing that: there exist $T > 0$ and $C > 0$ such that every solution of (7.1) satisfies

$$|(y_0, z_0, z_1)|_H^2 \leq C \int_0^T \int_{\Omega_1} |\nabla y|^2 dx dt \quad \forall (y_0, z_0, z_1) \in H. \quad (7.4)$$

Inequality (7.4) can be viewed as an observability estimate for equation (7.1) with observation on the heat subdomain. In principle, whether it holds or not depends very strongly on how the two components y and z of the solution are coupled along the interface. Indeed, the right-hand side term of (7.4) provides full information on y in Ω_1 and, consequently, also on the interface. Because of the continuity conditions on the interface this also yields information on z and its normal derivative on the interface. But how much of the energy of z we are able to obtain from this interface information has to be analyzed in detail. It depends on two facts. First, it may depend in a very significant way on whether the interface γ controls geometrically the wave domain Ω_2 or not. Second, of the trace of z (and its normal derivative) we recover.

REMARK 7.1. This argument also shows the close connections of the problems of control and that of the exponential decay of solutions of damped systems. Both end up being reducible to an observability inequality. This is particularly clear for the wave equation with localized damping

$$u_{tt} - \Delta u + \mathbb{1}_\omega u_t = 0.$$

In this case the energy is given by

$$E(t) = \frac{1}{2} \int [|u_t|^2 + |\nabla u|^2] dx,$$

and the energy dissipation law reads

$$\frac{d}{dt} E(t) = - \int_{\omega} |u_t|^2 dx.$$

The energy has an uniform exponential decay rate if and only if there exists some time T and constant $C > 0$ such that

$$E(0) \leq C \int_0^T \int_{\omega} |u_t|^2 dx dt.$$

Moreover, this observability estimate holds for the dissipative equation satisfied by u if and only if it holds for the conservative wave equation

$$\varphi_{tt} - \Delta \varphi = 0.$$

Thus we see that exponential decay is equivalent to observability which, as we know from previous sections, is also equivalent to controllability. This establishes a clear connection between controllability and stabilization. Here the argument has been developed for the wave equation but similar developments could be done for plate and Schrödinger equations and, more generally, for conservative evolution equations.

The fact that the exponential decay is equivalent to an observability inequality is also important for nonlinear problems. We refer to [170] and [44] for the analysis of the stabilization of nonlinear wave equations.

Returning to the coupled heat-wave system, as indicated in [166], there is no uniform decay for solutions of (7.1) even in one space dimension. The analysis in [166] exhibits the existence of a hyperbolic-like spectral branch such that the energy of the eigenvectors is concentrated in the wave domain and the eigenvalues have an asymptotically vanishing real part. This is obviously incompatible with the exponential decay rate. The approach in [166], based on spectral analysis, does not apply to multidimensional situations. But the 1- d result in [166] is a warning in the sense that one may not expect (7.4) to hold.

The exponential decay property also fails in several space dimensions, as the 1- d spectral analysis suggests. To prove this fact one has to build a family of solutions of the coupled system whose energy is mainly concentrated in the wave domain. This has been done in [169] following [146], using Gaussian beams [123,143], to construct approximate solutions of (7.1) which are highly concentrated along the generalized rays of the D'Alembert operator in the wave domain Ω_2 and are almost completely reflected on the interface γ . As we mentioned before, in the particular case of polygonal domains with a flat interface, one can do a simpler construction using plane waves.

This result on the lack of uniform exponential decay, which is valid for all geometric configurations, suggests that one can only expect a polynomial stability property of smooth solutions of (7.1) even under the geometric control condition, i.e., when the heat domain where the damping of the system is active is such that all rays of geometric optics propagating in the wave domain meet the interface in a uniform time. To prove this, we need to derive a weakened observability inequality. This can be done by viewing the whole system as a perturbation of the wave equation in the whole domain Ω , an argument that was introduced in [146] for the simpler interface conditions.

These results are summarized in the following section.

7.3. Decay properties

First of all, as mentioned above, solutions tend to zero as t goes to infinity but the decay rate is not exponential.

THEOREM 7.1. *For any given $(y_0, z_0, z_1) \in H$, the solution (y, z, z_t) of (7.1) tends to 0 strongly in H as $t \rightarrow \infty$, without any geometric assumption on the heat and wave domains other than Γ_2 being of positive capacity.*

But, at the contrary, there is no uniform exponential decay. In other words, the norm of the semigroup generated by the system $S(t)$ is one, $\|S(t)\| = 1$, as a linear and continuous operator from H to H , and this for all time $t > 0$.

To prove some decay rate it is convenient to view the whole coupled system as a perturbation of the wave equation in the union of the wave and heat domains. But for this method to work we need to assume that the heat domain Ω_1 satisfies the GCC. In this case the solutions of the wave equation in Ω

$$\begin{cases} \zeta_{tt} - \Delta \zeta = 0 & \text{in } \Omega \times (0, T), \\ \zeta = 0 & \text{on } \Gamma \times (0, T), \\ \zeta(0) = \zeta_0, \quad \zeta_t(0) = \zeta_1 & \text{in } \Omega \end{cases}$$

satisfy the following observability inequality (see [7])

$$\begin{aligned} & |\zeta_0|_{H_0^1(\Omega)}^2 + |\zeta_1|_{L^2(\Omega)}^2 \\ & \leq C \int_0^T \int_{\Omega_1} |\zeta_t|^2 dx dt \quad \forall (\zeta_0, \zeta_1) \in H_0^1(\Omega) \times L^2(\Omega), \end{aligned}$$

for T sufficiently large.

Under this condition the following holds.

THEOREM 7.2. *Assume that Ω_1 satisfies the GCC in Ω . Then there is a constant $C > 0$ such that for any $(y_0, z_0, z_1) \in D(\mathcal{A})$, the solution of (7.1) satisfies*

$$\|(y(t), z(t), z_t(t))\|_H \leq \frac{C}{t^{1/6}} \|(y_0, z_0, z_1)\|_{D(\mathcal{A})} \quad \forall t > 0. \quad (7.5)$$

REMARK 7.2. The domain of the generator of the semigroup of the coupled system is given by

$$D(\mathcal{A}) = \left\{ (Y_1, Y_2, Y_3) \in H \mid \Delta Y_1 \in L^2(\Omega_1), \Delta Y_2 \in L^2(\Omega_2), Y_3 \in H^1(\Omega_2), \right. \\ \left. Y_1|_{\Gamma_1} = Y_3|_{\Gamma_2} = 0, Y_1|_{\gamma} = Y_3|_{\gamma}, \right. \\ \left. \frac{\partial Y_1}{\partial \nu_1} \Big|_{\gamma} = -\frac{\partial Y_2}{\partial \nu_2} \Big|_{\gamma} \right\}.$$

As we have mentioned above it is not compactly embedded in H , except for the dimension $n = 1$.

REMARK 7.3. Theorem 7.2 is not sharp for $n = 1$ since, as proved in [166] using spectral analysis, the decay rate is $1/t^2$. Whether estimate (7.5) is sharp in several space dimensions is an open problem. However, its proof is rather rough and therefore it is very likely that it might be improved by a more subtle analysis of the interaction of the wave and heat components on the interface.

The proof of Theorem 7.2 is based on the following key weakened observability inequality for equation (7.1).

THEOREM 7.3. Assume that Ω_1 satisfies the GCC in Ω . Then there exist two constants T_0 and $C > 0$ such that for any $(y_0, z_0, z_1) \in D(\mathcal{A}^3)$, and any $T \geq T_0$, the solution of (7.1) satisfies

$$|(y_0, z_0, z_1)|_H \leq C |\nabla y|_{H^3(0,T; (L^2(\Omega_1))^n)}. \quad (7.6)$$

REMARK 7.4. Estimate (7.6) is a weakened observability inequality. Comparing it to (7.4), which is needed for the uniform exponential decay, we see that on the right-hand side term we are using a much stronger norm involving time derivatives up to order three. In order to get a better polynomial decay rate one should improve this observability estimate, using less time derivatives on the right-hand side term.

The main idea to prove Theorem 7.3 is as follows. Setting $w = y\chi_{\Omega_1} + z\chi_{\Omega_2}$, noting (7.1) and recalling that $\partial z_t / \partial \nu_2 = -\partial y_t / \partial \nu_1$ on $(0, T) \times \gamma$, and by $(y_0, z_0, z_1) \in D(\mathcal{A}^2)$, one sees that $w \in C([0, T]; H_0^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ satisfies

$$\begin{cases} w = (y_{tt} - y_t)\chi_{\Omega_1} + \left(\frac{\partial y}{\partial \nu_1} - \frac{\partial y_t}{\partial \nu_1}\right)\delta_{\gamma} & \text{in } (0, T) \times \Omega, \\ w = 0 & \text{on } (0, T) \times \Gamma, \\ w(0) = y_0\chi_{\Omega_1} + z_1\chi_{\Omega_2}, \quad w_t(0) = (\Delta y_0)\chi_{\Omega_1} + (\Delta z_0)\chi_{\Omega_2} & \text{in } \Omega. \end{cases} \quad (7.7)$$

Then, this weakened observability inequality holds from the GCC condition and energy estimates.

These results do not yield any decay rate in the case in which Ω_1 does not satisfy the GCC. This is for instance the case when Ω_1 is a convex subdomain in a convex domain Ω surrounded by the wave domain Ω_2 . In this case one expects a logarithmic decay rate. This was proved in [169] for the simpler interface conditions ($y = z$ instead of $y = z_t$). But the problem is open in the present case. This is once more due to the lack of compactness of the domain of the generator of the semigroup.

REMARK 7.5. The decay rates obtained in this section have been recently improved by Duyckaerts in [52] for C^∞ -domains, using sharp results on the interface interaction.

8. Some open problems

In this section we present some open problems related to the topics we have addressed in this chapter.

Spectral characterization of the controllability of the wave equation. In (3.15) we have presented a necessary and sufficient condition for the controllability and observability of the wave equation in terms of the uniform observability of certain eigenfunction packets. However, as indicated, this does not seem to give the sharp observability and controllability time. The problem of obtaining a spectral characterization of the observability property yielding the optimal time in several space dimensions is open.

Sharp observability estimates. In the context of the constant coefficient heat equation in Theorem 5.1 we referred to the sharp observability inequality

$$\sum_{k=1}^{\infty} |a_k|^2 e^{-c\sqrt{\lambda_k}} \leq C \int_0^T \int_{\omega} \varphi^2 \, dx \, dt \quad (8.1)$$

for the solutions φ of the adjoint heat equation (4.5). Note that the left-hand side of this inequality defines a norm of φ^0 that corresponds to the one in the domain of the operator $\exp(-c\sqrt{-\Delta})$. Characterizing the best constant c in this inequality in terms of the geometric properties of the domains Ω and ω is an open problem.

Obviously, the constant may also depend on the length of the time interval T . The problem may be made independent of T by considering the analogue in the infinite time. Indeed, for

$$\begin{cases} -\varphi_t - \Delta \varphi = 0 & \text{in } \Omega \times (-\infty, 0), \\ \varphi = 0 & \text{on } \partial\Omega \times (-\infty, 0), \\ \varphi(x, 0) = \varphi^0(x) & \text{in } \Omega, \end{cases} \quad (8.2)$$

the following inequality holds

$$\sum_{k=1}^{\infty} |a_k|^2 e^{-c\sqrt{\lambda_k}} \leq C \int_{-\infty}^0 \int_{\omega} \varphi^2 \, dx \, dt, \quad (8.3)$$

and the problem of determining the best constant $c > 0$ makes sense.

This problem is intimately related to the characterization of the best constant $A > 0$ in (5.2) for $a \equiv 0$, which is also an open problem. The problem of the best constant A in the inequality (5.2) can be also formulated independently of the length of the time interval T . For example, it is sufficient to consider the adjoint heat equation (8.2) in the infinite time $(-\infty, 0)$. The problem then consists in identifying the best constant $A > 0$ such that

$$\int_{-\infty}^0 \int_{\Omega} e^{-A/|t|} \varphi^2 dx dt \leq C \int_{-\infty}^0 \int_{\omega} \varphi^2 dx dt. \quad (8.4)$$

As we said above, by inspection of the proof of the inequality, one can get some rough estimates on A in terms of the weight function appearing in the Carleman inequality. But the obtention of sharp bounds or a complete characterization of the best constant is a fairly open problem. Let us briefly summarize what is known on it.

In [131] the problem was made independent of T differently, by taking the best asymptotic constant A as $T \rightarrow 0$. To be more precise, the following, uniform (in $0 \leq T \leq 1$) version was addressed:

$$\|\varphi(0)\|_{L^2(\Omega)}^2 \leq C e^{-\tilde{A}/T} \int_{-T}^0 \int_{\omega} \varphi^2(x, t) dx dt. \quad (8.5)$$

In (8.5) we look for a constants $\tilde{A}$ and C , uniform in $0 \leq T \leq 1$ for which (8.5) holds uniformly.

In that context explicit bounds on $\tilde{A}$ were obtained in [131]:

- The following lower bound on $\tilde{A}$ was proved:

$$\tilde{A} \geq \frac{\delta^2(\omega, \Omega)}{2}, \quad (8.6)$$

where $\delta(\omega, \Omega)$ stands for the largest geodesic distance between the set ω and any point in the domain Ω . This improves those previously obtained in [69] which referred to the radius of the largest ball included in $\Omega \setminus \omega$. This result (8.6) was proved in [131] using the well-known Varadhan's formula for the heat kernel in small time (see [160]).

Note that a similar lower bound on $\tilde{A}$ (or on A for any finite or infinite time-interval of the form $(-T, 0)$ with $T \leq \infty$) can be proved as a consequence of the upper Gaussian bounds on the heat kernel in [43], Theorem 3.2.7, p. 89.

- Upper bounds on $\tilde{A}$ were given in [131] in the case in which the geometric control condition is satisfied, using the transmutation method. This bound guarantees that

$$\tilde{A} \leq C \delta^2(\omega, \Omega), \quad (8.7)$$

for a suitable $C > 0$ which is independent of Ω and ω . This extends previous results in 1 – d by Seidman et al. [152] based on the use of Fourier series techniques.

The best constant $C > 0$ in this upper bound is unknown. In fact, it is unknown whether the lower bound (8.6) is optimal or not. In any case it seems hard to build

solutions of the heat equation that concentrate more than the Gaussian heat kernel, to improve the lower bound (8.6).

- The problem of getting the best bound is even open in 1- d . In [131] the estimate in [152] was significantly improved but the best constant is still unknown in this case too.

On the other hand, recently, in [135] it has been proved that the estimate (8.5) implies also (8.4) with the same constant which establishes a connection between both estimates.

But further investigation is needed for a complete characterization of the sharp values of A and/or $\tilde{A}$. In particular, as far as we know, even in 1- d there is no such a complete characterization.

We refer to [135] for the first results concerning the connection between the constants c and $\tilde{A}$ in (8.1) and (8.5).

The same can be said about the spectral estimate in (4.17). Characterizing the best constant B on the exponential degeneracy of the observability constant of finite linear combinations of eigenfunctions is an open problem. How this constant B is related to the best constants c and A in the inequalities above is an open problem.

Actually, as far as we know, there is no direct proof of the fact that the spectral observability inequality (4.17) implies the observability inequality for the heat equation. The existing proof is that due to Lebeau and Robbiano and passes through the property of null controllability and duality [108]. Moreover, the analysis in [134] shows that this method yields weaker observability estimates for the heat equation than those before since the observability constants obtained in this way have a stronger singularity at $T = 0$ of the order of e^{-C/T^β} for $\beta > 1$.

Observe also that the right-hand side term of inequality (8.3) can be written in Fourier series. Indeed, taking into account that $\varphi(x, t) = \sum_{j \geq 1} a_j \exp(\lambda_j t) \psi_j(x)$, we deduce that

$$\int_{-\infty}^0 \int_{\omega} \varphi^2 dx dt = \sum_{j, k \geq 1} a_j a_k \frac{\int_{\omega} \psi_j(x) \psi_k(x) dx}{\lambda_j + \lambda_k}. \quad (8.8)$$

Combining (8.3) and (8.8) we deduce that

$$\sum_{k=1}^{\infty} |a_k|^2 e^{-c\sqrt{\lambda_k}} \leq C \sum_{j, k \geq 1} a_j a_k \frac{\int_{\omega} \psi_j(x) \psi_k(x) dx}{\lambda_j + \lambda_k}. \quad (8.9)$$

Whether inequality (4.17) can be derived from (8.9) is an open problem.

Bang-bang controls. As described in Section 3.5, the problem of bang-bang controls for wave-like equations is still badly understood. In finite-dimensional dynamical systems and also for the heat equation, controls of minimal L^∞ -norm are of bang-bang form.

This is not the case for the wave equation as shown in [85] in 1- d using the D'Alembert formula to explicitly represent solutions. As explained in Section 3.5, in the context of the approximate controllability of the wave equation, quasibang-bang controls can be built both in one and several space dimensions. But a complete analysis of the actual structure

of these controls, which is intimately related with the structure of the nodal sets of the minimizers of the quadratic functionals for the adjoint system, is still to be done. The analysis in [84] shows that, even in 1- d , these quasibang-bang controls very rarely have an actual bang-bang structure. In the context of exact controllability the situation is even worse since relaxation phenomena occur and the minimizers of the corresponding minimization problem for the adjoint system may develop singular measures. The results in [84] also establish a clear relation between the lack of existence of bang-bang controls and the nonexistence of minimizers for the corresponding variational principle. The problem is totally unexplored in multi- d although, in view of the results in [84], obviously, one expects even a more complex picture.

A complete characterization of the set of data that are controllable by means of bang-bang controls has been obtained by Gugat and Leugering in [84] in 1- d using the explicit representation of solutions by means of D'Alembert's formula. But the problem is totally open in several space dimensions.

We have discussed the bang-bang principle in the following context. The time of control $T > 0$ and the initial and final data are given and we have considered the problem of finding bang-bang controls and relating this fact to the property of being of minimal L^∞ -norm within the class of admissible controls. Bang-bang controls arise also in the following alternate way. Assuming that controllability holds for some time T and considering initial and final data, and a given bound of the maximum size of the control allowed, one can look for the minimal time in which the system is controllable for those data under that control constraint. One expects the control to be of bang-bang form in that case too. This problem of minimal time control is well understood for finite-dimensional systems but much less is known in the context of PDE. We refer to the book [104] for the study of this problem for 1- d wave and heat processes.

Control of semilinear heat equations. As we mentioned above, in [70] it was proved that semilinear heat equations are null controllable in an arbitrarily small time and from any open subset ω of the domain Ω , for nonlinearities that, at infinity, grow slower than $s \log^{3/2}(s)$. On the other hand, there are examples in which, because of blow-up phenomena, this result fails to be true when the nonlinearity grows as $s \log^p(s)$ with $p > 2$. The optimality results of Section 5.2 show that one can not expect the classical method (based on linearization, Carleman inequalities and fixed points) to work in the range $3/2 < p \leq 2$. Whether null controllability holds in an arbitrarily small time in that range of nonlinearities is an interesting open problem.

For power-like nonlinearities, or even for nonlinearities growing at infinity as $s \log^p(s)$ with $p > 2$, null controllability may fail because of blow up [70]. On the other hand, for nonlinearities with the good sign it is also well known that for null controllability to hold the time of control has to be taken large enough, depending on the size of the initial datum to be controlled [5]. Recently, it has also been proved that, despite these possible nonlinear effects, these equations are controllable in the sense that two stationary solutions in the same connected component (within the space of stationary solutions), can be connected by means of a suitable control (see [38]).

Determining what the situation is for the Navier–Stokes equations constitutes an open problem. So far the existing null controllability results are local and need the time to be

large when the initial data are large. Whether this is necessary or not is an open problem. In [65] it has been recently proved in the context of the 1- d Burgers equation that the time of control actually needs to be large when the initial data are large. But the multi- d analogue for the Navier–Stokes equations is open, as far as we know.

Degenerate parabolic problems. Heat equations with degenerate coefficients in the principal part and possible applications to nonlinear parabolic equations as the porous medium or the p -Laplacian one, for instance, is a widely open subject of research. We refer to [24] for the first results in this direction, concerning the 1- d linear heat equation with a space-dependent coefficient degenerating on a single point, in a polynomial way.

Equations with low regularity coefficients. To generalize the eigenfunction estimate in Theorem 4.16 for elliptic operators and systems with low regularity coefficients and in nonsmooth domains is an open problem. That would allow adapting Lebeau and Robbiano's [108] strategy for proving the null controllability of the underlying parabolic equation/system. So far, the technique in [108], based on Carleman inequalities, only applies for smooth coefficients and domains.

As mentioned in Section 6, Carleman inequalities can be directly applied to obtain observability inequalities for heat equations with variable coefficients in the principal part which are a small L^∞ perturbation of constant coefficients. Furthermore, the smallness condition depends on time. Whether these smallness conditions are necessary or not is an open problem. The same can be said about the observability estimates for elliptic equations addressed in Section 6.

As far as we know, there is no result in the literature showing the lack of null controllability of the heat equation with bounded measurable coefficients. This is an interesting and possibly difficult open problem. The results in [1] show that, in 1- d , the same is true for heat equations with bounded measurable coefficients.

Spectral characterization of controllability of parabolic equations. In the context of the wave equation we have given a spectral necessary and sufficient condition (3.15) for controllability in terms of the observability of wave packets, combination of eigenfunctions corresponding to nearby eigenfunctions. A similar characterization is unknown for the heat equation and, more generally, for time-irreversible systems.

The iterative method in [108] uses the eigenfunction estimate (4.17) in Theorem 4.16 but does not suffice in itself to yield the observability estimate for the heat equation directly. The argument in [108] passes through the property of null controllability of the controlled system and, as a corollary, gives the observability property. It would be interesting to develop a direct iterative argument showing that the eigenfunction estimate (4.17) implies the observability inequality for the heat equation.

The spectral estimate also implies the controllability and observability of fractional order equations of the form (4.22) with $\alpha > 1/2$. Thus, a complete characterization of (4.17) in terms of the associated evolution problems is still to be found. It is however important to note that the approach based on eigenfunction estimates of the form Theorem 4.1 is certainly more limited than that consisting in addressing directly the heat equation by Car-

leman inequalities, which allows, for instance, considering heat equations with coefficients depending both on space x and time t , as in Section 6.

Systems of parabolic equations. Most of the literature on the null control of parabolic equations refers to the scalar case. A lot remains to be done to better understand the null controllability of systems. In view of the results we have described for the scalar heat equation one could expect that parabolic systems share the same property of being null controllable in any time and from any open nonempty subset. But systems also have the added possibility of controlling the whole state with less components of controls than the state of the system has. In fact, a very natural condition for that to hold is that the underlying algebraic structure of the system satisfies the Kalman rank condition. Some preliminary results in that direction can be found in [153], [81] and [83] but a complete answer is still to be found.

Another source of problems in which Carleman and observability inequalities for systems arise is that of the insensitizing control. This notion was introduced in the infinite-dimensional context by Lions in [114], the goal being to reduce the sensitivity of partial measurements of solutions of a given equation with respect to perturbations (of initial data, right hand side terms, etc.). As pointed out by Lions, in an appropriate setting, this amounts to prove an observability estimate (or unique continuation property if the notion of insensitivity is relaxed to some approximate insensitivity one) for a cascade system in which the equations are coupled in a diagonal way, but with the very peculiarity that one is forward in time, the other one being a backward equation. Interesting results in this direction have been proved in [12,45,63], in particular. Still, a complete theory is to be developed, in particular in the case where the control acts on a region with empty intersection with the subset of the domain to be insensitized (the first results in that direction for the multidimensional heat equation, using Fourier series techniques have been recently obtained by [47]).

Control and homogenization. In this chapter we have not addressed the issue of homogenization in controllability, and more generally, that on the behavior of the controllability properties under singular perturbations of the system under consideration. These problems can be formulated for a wide class of equations. In particular, in the context of homogenization of PDE with rapidly oscillating coefficients it consists roughly in analyzing whether the controls converge to the control of the limit homogenized equation as the frequency of oscillation of the coefficients tends to infinity, or the microstructure gets finer and finer. The problem is relevant in applications. Indeed, when the convergence of controls holds, one can use the control of the homogeneous limit homogenized medium (which is much easier to compute because of the lack of heterogeneities) to control the heterogeneous equation. This property holds often in the context of approximate controllability, but rarely does for exact or null controllability problems.

We now summarize the existing results and some open problems in this field. We refer to [27] for further developments in this direction and for a complete discussion of this issue.

(1) The controls have been proved to converge in the context of 1- d the heat equation with rapidly oscillating periodic coefficients [122]. The multidimensional analogue constitutes an interesting open problem.

(2) In the frame of approximate controllability, the convergence of controls was proved for the heat equation with rapidly oscillating coefficients in [172]. The same proof applies to a wide class of PDE. It is sufficient to apply the variational techniques we have developed to construct approximate controls, combined with Γ -convergence arguments [42]. Despite this fact, addressing exact or null controllability problems is of much greater complexity since it requires of observability estimates which are independent of the vanishing parameter measuring the size of the microstructure.

(3) In the context of the wave equation the property of exact controllability fails to be uniform as the microstructure gets finer and finer. This is due to resonance phenomena of high frequency waves for which the wavelength is of the order of the microstructure. However, the controls can be proved to be uniformly bounded, and one can pass to the limit to get the control of the homogenized equation, when one relaxes the controllability condition to that of controlling only the low frequencies. The controlled low frequencies are precisely those that avoid resonance with the microstructure. We refer to [27] for a survey article on this topic. This result was first proved by Castro and Zuazua in 1- d in [25] and later extended to the multidimensional case by Lebeau [107] using Bloch waves decompositions and microlocal analysis techniques.

In view of this and using the methods of transforming control results for wave equations into control results for heat equations, one can show that the heat equation with rapidly oscillating coefficients is uniformly partially null controllable when the control subdomain satisfies the GCC for the underlying homogenized problem. Here partial null controllability refers to the possibility of controlling to zero the projection of solutions on the low-frequency components avoiding resonances with the microstructure. Getting the uniform null controllability out of this is an open problem. The three-step method developed in [121] and [122] could be applied if we had a Carleman inequality for the multi- d heat equation with rapidly oscillating coefficients with an observability constant of the form $\exp(C/\varepsilon^\alpha)$ with $\alpha < 2$. But such an estimate is unknown. The direct application of the existing Carleman inequalities provides an estimate of the form $\exp(\exp(C/\varepsilon))$ which is far from being sufficient.

Another drawback of this result is that for it to hold one needs to impose a GCC on the control subset, a fact whose necessity has not been justified in the context of parabolic equations.

(4) Similar problems arise in the context of perforated domains. One may consider, for instance, the case of periodically perforated domains with small holes in the sense of Cioranescu and Murat [31]. In that case, for Dirichlet boundary conditions, the limit effective equation is the wave equation itself in the whole domain, with possibly a lower-order perturbation. In that situation it is known that the exact controllability property passes to the limit provided the controls are applied everywhere in the boundary: the external one and that of the holes (see [30]). The question of whether by filtering the high frequencies one can achieve the uniform controllability from the exterior boundary, as for rapidly oscillating coefficients in [107], is open.

(5) The same can be said about the null controllability of the heat equation.

Fluid–structure interaction. Important work also remains to be done in the context of the models for fluid–structure interaction we have discussed in Section 7. We include here a brief description of some of the most relevant ones.

- Logarithmic decay without the GCC. Inspired on [147], it seems natural to expect a logarithmic decay result for system (7.1) without the GCC. This has been done successfully in [146] when replacing the interface condition $y = z_t$ by $y = z$. However, there is a difficulty when addressing the interface condition $y = z_t$ directly which is related to the lack of compactness for (7.1) in multidimensions.
- More complex and realistic models. The model under consideration would be more realistic replacing the wave equation in system (7.1) by the system of elasticity and the heat equation by the Stokes system, and the fluid–solid interface by a free boundary. It would be interesting to extend the results in Section 7 to these situations but this remains to be done.

The problem would be even more realistic when considering nonlinear equations, as for instance the Navier–Stokes equations for the fluid. But, to the best of our knowledge, very little is known about the well-posedness and the long time behavior for the solutions to the corresponding equations. (We refer to [13] and [14] for some existence results of weak solutions and [39] for local smooth solutions in 3- d .)

- Control problems. In [166], we analyzed the null controllability problem for system (7.1) in one space dimension by means of spectral methods. It was found that the controllability results depend strongly on whether the control enters the system through the wave component or the heat one. When the control acts on the boundary of the wave interval one obtains null controllability in the energy space. However, when the control acts of the boundary of the heat interval, null controllability holds in a much smaller space. This is due to the existence of an infinite branch of eigenfunctions that are weakly dissipated and strongly concentrated on the wave interval. Therefore, the control affects these spectral components exponentially weakly at high frequencies. Because of this the initial data to be controlled need to have exponentially small Fourier coefficients on that spectral branch. This problem is completely open in several space dimensions. Two different situations need to be considered also in the multidimensional case depending on whether the control acts on a subset of the wave or heat domain. The answer to the problem may also depend on whether the set in which the control enters controls geometrically the whole domain or not. In 1- d this condition is automatically satisfied because the only possible rays are segments that cover the whole domain under propagation.

Controllability of stochastic PDE. Extending the results we have presented in this chapter to the stochastic framework is a widely open subject of research. There are however some interesting results in this direction. We refer for instance to the article [156] where Carleman inequalities were proved for stochastic parabolic equations.

Acknowledgements

First of all, I would like to thank Constantine Dafermos and Eduard Feireisl for having thought on me for writing this chapter and having given me the opportunity to do it. This

chapter is somehow a summary of part of the work I have done in this field in close collaboration with my colleagues, coworkers and Ph.D. students. I am grateful to all of them. In particular, I would like to express my gratitude to J.-L. Lions with whom I got initiated in this subject and to R. Glowinski who, later, played the same role in what concerns the Numerical Analysis aspects. My thanks go also to J.-P. Puel for his continuous encouragement and for so many fruitful discussions during so many years that have been extremely influential on the formulation and solving of most of the problems addressed in these notes. Finally, I would like also to mention some of my colleagues who with I had the opportunity to develop part of the theory and learn many many things and, in particular, C. Castro, E. Fernández-Cara, G. Lebeau, S. Micu and X. Zhang. I would also like to thank G. Leugering and M. Gugat for interesting and useful discussions on bang-bang controls, L. Escauriaza for his advice on unique continuation problems and C. Palencia for his comments on Gaussian bounds for heat kernels. Finally, I thank also E. Fernández-Cara, S. Guerrero, L. Miller, E. Trélat and X. Zhang for their useful comments on the first version of this paper that allowed to improve its presentation and to avoid some inaccuracies.

Supported by Grant MTM2005-00714 of the Spanish MEC, the DOMINO Project CIT-370200-2005-10 in the PROFIT program of the MEC (Spain), the SIMUMAT project S-0505/ESP/0158 of the CAM (Spain) and by the European network “Smart Systems”.

References

- [1] G. Alessandrini and L. Escauriaza, *Null-controllability of one-dimensional parabolic equations*, ESAIM Control Optim. Calc. Var., to appear.
- [2] S. Alinhac, *Non unicité du problème de Cauchy*, Ann. of Math. **117** (1983), 77–108.
- [3] S. Angenent, *The zero set of a solution of a parabolic equation*, J. Reine Angew. Math. **390** (1988), 76–96.
- [4] S. Anita and V. Barbu, *Null controllability of nonlinear convective heat equations*, ESAIM Control Optim. Calc. Var. **5** (2000), 157–173.
- [5] S. Anita and D. Tataru, *Null controllability for the dissipative semilinear heat equation*, Appl. Math. Optim. **46** (2/3) (2002), 97–105. Special issue dedicated to the memory of Jacques-Louis Lions.
- [6] S.A. Avdonin and S.A. Ivanov, *Families of Exponentials. The Method of Moments in Controllability Problems for Distributed Parameter Systems*, Cambridge Univ. Press (1995).
- [7] C. Bardos, G. Lebeau and J. Rauch, *Sharp sufficient conditions for the observation, control and stabilization of waves from the boundary*, SIAM J. Control. Optim. **30** (1992), 1024–1065.
- [8] K. Beauchard, *Local controllability of a 1-D Schrödinger equation*, J. Math. Pures Appl. **84** (2005), 851–956.
- [9] K. Beauchard and J.M. Coron, *Controllability of a quantum particle in a moving potential well*, J. Funct. Anal. **232** (2) (2006), 328–389.
- [10] L. Bers, F. John and M. Schechter, *Partial Differential Equations*, Interscience, New York (1964).
- [11] L. Bers and L. Nirenberg, *On a representation theorem for linear elliptic systems with discontinuous coefficients and its applications*, Convegno Internazionale sulle Equazioni alle Derivate Parziali, Cremonese, Roma (1955), 111–138.
- [12] O. Bodart, M. González-Burgos and R. Pérez-García, *Existence of insensitizing controls for a semilinear heat equation with a superlinear nonlinearity*, Comm. Partial Differential Equations **29** (7/8) (2004), 1017–1050.
- [13] M. Boulakia, *Existence of weak solutions for the motion of an elastic structure in an incompressible viscous fluid*, C. R. Math. Acad. Sci. Paris **336** (12) (2003), 985–990.
- [14] M. Boulakia, *Existence of weak solutions for an interaction problem between an elastic structure and a compressible viscous fluid*, C. R. Math. Acad. Sci. Paris **340** (2) (2005), 113–118.

- [15] M. Boulakia and A. Osses, *Two-dimensional local null controllability of a right structure in a Navier–Stokes fluid*, C. R. Math. Acad. Sci. Paris Ser I **343** (2) (2006), 105–109.
- [16] N. Burq, *Contrôle de l'équation des ondes dans des ouverts peu réguliers*, Asympt. Anal. **14** (1997), 157–191.
- [17] N. Burq, *Contrôle de l'équation des ondes dans des ouverts comportant des coins*, Bull. Soc. Math. France **126** (4) (1998), 601–637. (With an appendix in collaboration with J.-M. Schlenker.)
- [18] N. Burq, *Décroissance de l'énergie locale de equation des ondes pour le problème extérieur et absence de résonance au voisinage du réel*, Acta Math. **180** (1998), 1–29.
- [19] N. Burq and P. Gérard, *Condition nécessaire et suffisante pour la contrôlabilité exacte des ondes*, C. R. Acad. Sci. Paris **325** (1997), 749–752.
- [20] N. Burq and G. Lebeau, *Mesures de défaut de compacité, application au système de Lamé*, Ann. Sci. École Norm. Sup. **34** (6) (2001), 817–870.
- [21] N. Burq and M. Zworski, *Geometric control in the presence of a black box*, J. Amer. Math. Soc. **17** (2) (2004), 443–471.
- [22] V. Cabanillas, S. de Menezes and E. Zuazua, *Null controllability in unbounded domains for the semilinear heat equation with nonlinearities involving gradient terms*, J. Optim. Theory Appl. **110** (2) (2001), 245–264.
- [23] P. Cannarsa, P. Martinez and J. Vancostenoble, *Null controllability of the heat equation in unbounded domains by a finite measure control region*, ESAIM Control Optim. Calc. Var. **10** (3) (2004), 381–408.
- [24] P. Cannarsa, P. Martinez and J. Vancostenoble, *Null controllability of degenerate heat equations*, Adv. Differential Equations **10** (2) (2005), 153–190.
- [25] C. Castro and E. Zuazua, *Low frequency asymptotic analysis of a string with rapidly oscillating density*, SIAM J. Appl. Math. **60** (2000), 1205–1233.
- [26] C. Castro and E. Zuazua, *Concentration and lack of observability of waves in highly heterogeneous media*, Arch. Ration. Mech. Anal. **164** (1) (2002), 39–72.
- [27] C. Castro and E. Zuazua, *Control and homogenization of wave equations*, Homogenization 2001. Proc. First HMS2000 Internat. School and Conf. Homogenization, L. Carbone and R. De Arcangelis eds, GAKUTO Internat. Ser. Math. Sci. Appl., Vol. 18, Gakkōtoshō, Tokyo (2002), 45–94.
- [28] D. Chae, O.Yu. Imanuvilov and S.M. Kim, *Exact controllability for semilinear parabolic equations with Neumann boundary conditions*, J. Dynam. Control Systems **2** (1996), 449–483.
- [29] P. Chen and H. Qin, *Controllability of linear systems in Banach spaces*, Systems Control Lett. **45** (2002), 155–161.
- [30] D. Cioranescu, P. Donato and E. Zuazua, *Exact boundary controllability for the wave equation in domains with small holes*, J. Math. Pures Appl. **71** (1992), 343–377.
- [31] D. Cioranescu and F. Murat, *A strange term coming from nowhere*, Topics in the Mathematical Modelling of Composite Materials, Progr. Nonlinear Differential Equations Appl., Vol. 31, Birkhäuser, Boston, MA, (1997), 45–93.
- [32] J.-M. Coron, *Global asymptotic stabilization for controllable systems without drift*, Math. Control Signals Systems **5** (3) (1992), 295–312.
- [33] J.-M. Coron, *On the controllability of 2-D incompressible perfect fluids*, J. Math. Pures Appl. **75** (2) (1996), 155–188.
- [34] J.-M. Coron, *On the controllability of the 2-D incompressible Navier–Stokes equations with the Navier slip boundary conditions*, ESAIM Control Optim. Calc. Var. **1** (1996), 35–75.
- [35] J.-M. Coron, *Return method: Some applications to flow control*, Mathematical Control Theory, Part 1, 2 (Trieste, 2001), ICTP Lect. Notes, VIII, Abdus Salam Int. Cent. Theoret. Phys., Trieste (2002), 655–704.
- [36] J.-M. Coron and A. Fursikov, *Global exact controllability of the 2D Navier–Stokes equations on a manifold without boundary*, Russian J. Math. Phys. **4** (4) (1996), 1–19.
- [37] J.-M. Coron and S. Guerrero, *Singular optimal control: A linear 1-D parabolic–hyperbolic example*, Asymptot. Anal. **44** (3/4) (2005), 237–257.
- [38] J.-M. Coron and E. Trélat, *Global steady-state controllability of one-dimensional semilinear heat equations*, SIAM J. Control Optim. **43** (2) (2004), 549–569.
- [39] D. Coutand and S. Shkoller, *The interaction between elastodynamics and the Navier–Stokes equations*, Arch. Ration. Mech. Anal. **179** (3) (2006), 303–352.

- [40] S. Cox and E. Zuazua, *The rate at which energy decays in the string damped at one end*, Indiana Univ. Math. J. **44** (2) (1995), 545–573.
- [41] R. Dáger and E. Zuazua, *Wave Propagation, Observation and Control in 1-d Flexible Multistructures*, Math. Appl., Vol. 50, Springer-Verlag (2006).
- [42] G. Dal Maso, *An Introduction to Γ -convergence*, Birkhäuser, Boston, MA (1993).
- [43] E.B. Davies, *Heat Kernels and Spectral Theory*, Cambridge Tracts in Math., Vol. 92, Cambridge Univ. Press (1989).
- [44] B. Dehman, G. Lebeau and E. Zuazua, *Stabilization and control of the semilinear subcritical wave equation*, Ann. Sci. Ecole Norm. Sup. **36** (4) (2003), 525–551.
- [45] L. de Teresa, *Controls insensitizing the semilinear heat equation*, Comm. Partial Differential Equations **25** (1/2) (2000), 39–72.
- [46] L. de Teresa, *Approximate controllability of a semilinear heat equation in $\mathbb{R}^N$* , SIAM J. Control. Optim. **36** (6) (1998), 2128–2147.
- [47] L. de Teresa and O. Kavian, *Unique continuation principle for systems of parabolic equations*, Preprint (2005).
- [48] L. de Teresa and E. Zuazua, *Approximate controllability of the heat equation in unbounded domains*, Nonlinear Anal. **37** (8) (1999), 1059–1090.
- [49] A. Doubova and E. Fernández-Cara, *Some control results for simplified one-dimensional models of fluid–solid interaction*, Math. Models Methods Appl. Sci. **15** (5) (2005), 783–824.
- [50] A. Doubova, E. Fernández-Cara, M. González-Burgos and E. Zuazua, *On the controllability of parabolic systems with a nonlinear term involving the state and the gradient*, SIAM J. Control Optim. **41** (3) (2002), 798–819.
- [51] A. Doubova, A. Osses and J.-P. Puel, *Exact controllability to trajectories for semilinear heat equations with discontinuous diffusion coefficients*, A tribute to J.-L. Lions, ESAIM Control Optim. Calc. Var. **8** (2002), 621–661.
- [52] Th. Duyckaerts, *Optimal decay rates of the energy of an hyperbolic–parabolic system coupled by an interface*, Preprint (2005).
- [53] Th. Duyckaerts, X. Zhang and E. Zuazua, *On the optimality of the observability inequalities for parabolic and hyperbolic systems with potentials*, Ann. Inst. H. Poincaré Anal. Non Linéaire, to appear.
- [54] Y. Ekeland and R. Temam, *Analyse convexe et problèmes variationnelles*, Dunod, Gauthier-Villars, Paris (1974).
- [55] L. Escauriaza, G. Seregin and V. Sverák, *Backward uniqueness for parabolic equations*, Arch. Ration. Mech. Anal. **169** (1) (2003), 147–157.
- [56] C. Fabre, J.-P. Puel and E. Zuazua, *Contrôlabilité approchée de l’équation de la chaleur linéaire avec des contrôles de norme L^∞ minimale*, C. R. Acad. Sci. Paris **316** (1993), 679–684.
- [57] C. Fabre, J.-P. Puel and E. Zuazua, *Approximate controllability of the semilinear heat equation*, Proc. Royal Soc. Edinburgh Sect. A **125** (1995), 31–61.
- [58] C. Fabre, J.-P. Puel and E. Zuazua, *On the density of the range of the semigroup for semilinear heat equations*, Control and Optimal Design of Distributed Parameter Systems, IMA Vol. Math. Appl., Vol. 70, Springer-Verlag, New York (1995), 73–92.
- [59] H.O. Fattorini, *Infinite Dimensional Optimization and Control Theory*, Encyclopedia of Mathematics and Its Applications, Vol. 62, Cambridge Univ. Press (1999).
- [60] H. Fattorini and D.L. Russell, *Exact controllability theorems for linear parabolic equations in one space dimension*, Arch. Ration. Mech. Anal. **43** (1971), 272–292.
- [61] H. Fattorini and D.L. Russell, *Uniform bounds on biorthogonal functions for real exponentials with an application to the control theory of parabolic equations*, Quart. Appl. Math. **32** (1974), 45–69.
- [62] E. Feireisl and E. Zuazua, *Global attractors for semilinear wave equations with locally distributed damping and critical exponent*, Comm. Partial Differential Equations **18** (9/10) (1993), 1539–1556.
- [63] E. Fernández-Cara, G.C. Garcia and A. Osses, *Controls insensitizing the observation of a quasi-geostrophic ocean model*, SIAM J. Control Optim. **43** (5) (2005), 1616–1639.
- [64] E. Fernández-Cara, M. González-Burgos, S. Guerrero and J.-P. Puel, *Null controllability of the heat equation with Fourier boundary conditions: The linear case*, ESAIM Control Optim. Calc. Var. **12** (3) (2006), 442–466.

- [65] E. Fernández-Cara and S. Guerrero, *Remarks on the null controllability of the Burgers equation*, C. R. Acad. Sci. Paris **341** (2005), 229–232.
- [66] E. Fernández-Cara and S. Guerrero, *Global Carleman estimates for solutions of parabolic systems defined by transposition and some applications to controllability*, Appl. Math. Res. Express **2006** (2006), article ID 75090, 31 pp.
- [67] E. Fernández-Cara and S. Guerrero, *Global Carleman inequalities for parabolic systems and applications to controllability*, SIAM J. Control Optim., to appear.
- [68] E. Fernández-Cara, S. Guerrero, O.Yu. Imanuvilov and J.-P. Puel, *Local exact controllability of the Navier–Stokes system*, J. Math. Pures Appl. **83** (2004), 1501–1542.
- [69] E. Fernández-Cara and E. Zuazua, *The cost of approximate controllability for heat equations: The linear case*, Adv. Differential Equations **5** (4/6) (2000), 465–514.
- [70] E. Fernández-Cara and E. Zuazua, *Null and approximate controllability for weakly blowing-up semilinear heat equations*, Ann. Inst. H. Poincaré Anal. Non Linéaire **17** (5) (2000), 583–616.
- [71] E. Fernández-Cara and E. Zuazua, *On the null controllability of the one-dimensional heat equation with BV coefficients*, Comput. Appl. Math. **21** (1) (2002), 167–190.
- [72] X. Fu, *A weighted identity for partial differential operators of second order and applications*, C. R. Acad. Sci. Paris **342** (2006), 579–584.
- [73] A.V. Fursikov, *Exact boundary zero controllability of three dimensional Navier–Stokes equations*, J. Dynam. Control Systems **1** (3) (1995), 325–350.
- [74] A.V. Fursikov, *Optimal Control of Distributed Systems. Theory and Applications*, Transl. Math. Monogr., Vol. 187, Amer. Math. Soc., Providence, RI (2000).
- [75] A.V. Fursikov and O.Yu. Imanuvilov, *On exact boundary zero-controllability of two-dimensional Navier–Stokes equations*, Acta Appl. Math. **37** (1994), 67–76.
- [76] A.V. Fursikov and O.Yu. Imanuvilov, *Controllability of Evolution Equations*, Lecture Notes Ser., Vol. 34, Seoul Nat. Univ., Seoul (1996).
- [77] A. Fursikov and O.Yu. Imanuvilov, *Exact controllability of the Navier–Stokes and Boussinesq equations*, Russian Math. Surveys **54** (3) (1999), 565–618.
- [78] P. Gérard, *Microlocal defect measures*, Comm. Partial Differential Equations **16** (1991), 1761–1794.
- [79] J.-M. Ghidaglia, *Some backward uniqueness results*, Nonlinear Anal. **10** (1986), 777–790.
- [80] O. Glass, *Exact boundary controllability of 3-D Euler equation*, ESAIM Control Optim. Calc. Var. **5** (2000), 1–44.
- [81] M. González-Burgos and R. Pérez-García, *Controllability results for some nonlinear coupled parabolic systems by one control force*, Asymptot. Anal. **46** (2) (2006), 123–162.
- [82] P. Grisvard, *Contrôlabilité exacte des solutions de l'équation des ondes en présence de singularités*, J. Math. Pures Appl. **68** (1989), 215–259.
- [83] S. Guerrero, *Null controllability of some systems of two parabolic equations with one control force*, Preprint (2005).
- [84] M. Gugat and G. Leugering, *L^∞ norm minimal control of the wave equation: On the weakness of the bang-bang principle*, Preprint (2005); ESAIM Control Optim. Calc. Var., to appear.
- [85] M. Gugat, G. Leugering and G. Sklyar, *L^p -optimal boundary control for the wave equation*, SIAM J. Control Optim. **44** (2005), 49–74.
- [86] S. Hansen and E. Zuazua, *Controllability and stabilization of strings with point masses*, SIAM J. Control Optim. **33** (5) (1995), 1357–1391.
- [87] A. Haraux, *Systèmes dynamiques dissipatifs et applications*, RMA, Vol. 17, Masson, Paris (1990).
- [88] A. Haraux and S. Jaffard, *Pointwise and spectral controllability for plate vibrations*, Rev. Mat. Iberoamericana **7** (1) (1991), 1–24.
- [89] A. Haraux and E. Zuazua, *Decay estimates for some semilinear damped hyperbolic problems*, Arch. Ration. Mech. Anal. **100** (2) (1988), 191–206.
- [90] T.I. Hesla, *Collisions of smooth bodies in viscous fluids: A mathematical investigation*, PhD Thesis, Univ. Minnesota (2005).
- [91] M. Hillairet, *Lack of collision between solid bodies in a 2d incompressible viscous flow*, Preprint (2006).
- [92] L.F. Ho, *Observabilité frontière de l'équation des ondes*, C. R. Acad. Sci. Paris **302** (1986), 443–446.
- [93] L. Hörmander, *Linear Partial Differential Equations*, Springer-Verlag (1969).

- [94] O.Yu. Imanuvilov, *Boundary controllability of parabolic equations*, Russian Acad. Sci. Sb. Math. **186** (1995), 109–132 (in Russian).
- [95] O.Yu. Imanuvilov, *In Carleman estimates for hyperbolic equations*, Asymptot. Anal. **32** (2002), 185–220.
- [96] O.Yu. Imanuvilov and J.-P. Puel, *Global Carleman estimates for weak solutions of elliptic nonhomogeneous Dirichlet problems*, C. R. Acad. Sci. Paris **335** (2002), 33–38.
- [97] O.Yu. Imanuvilov and J.-P. Puel, *Global Carleman estimates for weak solutions of elliptic nonhomogeneous Dirichlet problems*, IMRN **16** (2003), 883–913.
- [98] O.Yu. Imanuvilov and M. Yamamoto, *Carleman estimate for a parabolic equation in Sobolev spaces of negative order and its applications*, Control of Nonlinear Distributed Parameter Systems, G. Chen et al., eds, Marcel-Dekker (2000), 113–137.
- [99] A.E. Ingham, *Some trigonometric inequalities with applications to the theory of series*, Math. Z. **41** (1936), 367–379.
- [100] V. Isakov, *Inverse Problems for Partial Differential Equations*, Springer-Verlag, Berlin (1988).
- [101] V. Komornik, *Exact Controllability and Stabilization, The Multiplier Method*, Wiley, Chichester/Masson, Paris (1994).
- [102] V. Komornik and P. Loret, *Fourier Series in Control Theory*, Springer Monogr. Math., Springer-Verlag, New York (2005).
- [103] V. Komornik and E. Zuazua, *A direct method for the boundary stabilization of the wave equation*, J. Math. Pures Appl. **69** (1) (1990), 33–55.
- [104] W. Krabs, *On Moment Theory and Controllability of One-Dimensional Vibrating Systems and Heating Processes*, Lecture Notes in Control and Inform. Sci., Vol. 173, Springer-Verlag, New-York (1992).
- [105] I. Lasiecka and R. Triggiani, *Control Theory for Partial Differential Equations: Continuous and Approximation Theories, Vol. I*, Cambridge Univ. Press (2000).
- [106] G. Lebeau, *Contrôle analytique I: Estimations a priori*, Duke Math. J. **68** (1) (1992), 1–30.
- [107] G. Lebeau, *The wave equation with oscillating density: Observability at low frequency*, ESAIM Control Optim. Calc. Var. **5** (2000), 219–258.
- [108] G. Lebeau and L. Robbiano, *Contrôle exact de l'équation de la chaleur*, Comm. Partial Differential Equations, **20** (1995), 335–356.
- [109] G. Lebeau and E. Zuazua, *Null controllability of a system of linear thermoelasticity*, Arch. Ration. Mech. Anal. **141** (4) (1998), 297–329.
- [110] G. Lebeau and E. Zuazua, *Decay rates for the linear system of three-dimensional system of thermoelasticity*, Arch. Ration. Mech. Anal. **148** (1999), 179–231.
- [111] E.B. Lee and L. Markus, *Foundations of Optimal Control Theory*, SIAM Ser. Appl. Math., Wiley, New York (1967).
- [112] J.-L. Lions, *Contrôlabilité exacte, stabilisation et perturbations de systèmes distribués*, Tome 1. Contrôlabilité Exacte, RMA, Vol. 8, Masson, Paris (1988).
- [113] J.-L. Lions, *Exact controllability, stabilizability and perturbations for distributed systems*, SIAM Rev. **30** (1988), 1–68.
- [114] J.L. Lions, *Remarques préliminaires sur le contrôle des systèmes à données incomplètes*, Actas del Congreso de Ecuaciones Diferenciales y Aplicaciones (CEDYA), Universidad de Málaga (1989), 43–54.
- [115] J.-L. Lions, *Remarks on approximate controllability*, J. Anal. Math. **59** (1992), 103–116.
- [116] J.-L. Lions and B. Malgrange, *Sur l'unicité rétrograde dans les problèmes mixtes paraboliques*, Math. Scand. **8** (1960), 277–286.
- [117] K. Liu, *Locally distributed control and damping for the conservative systems*, SIAM J. Control. Optim. **35** (5) (1997), 1574–1590.
- [118] K. Liu, Z. Liu and B. Rao, *Exponential stability of an abstract nondissipative linear system*, SIAM J. Control Optim. **40** (1) (2001), 149–165.
- [119] W.J. Liu and E. Zuazua, *Decay rates for dissipative wave equations*, Ricerche Mat. **XLVIII** (1999), 61–75.
- [120] Z. Liu and B. Rao, *Stabilité exponentielle des équations des ondes avec amortissement local de Kelvin-Voigt*, C. R. Math. Acad. Sci. Paris **339** (11) (2004), 769–774.
- [121] A. López, X. Zhang and E. Zuazua, *Null controllability of the heat equation as a singular limit of the exact controllability of dissipative wave equations*, J. Math. Pures Appl. **79** (2000), 741–809.
- [122] A. López and E. Zuazua, *Uniform null-controllability for the one-dimensional heat equation with rapidly oscillating coefficients*, C. R. Acad. Sci. Paris **326** (1998), 955–960.

- [123] F. Macià and E. Zuazua, *On the lack of observability for wave equations: A Gaussian beam approach*, Asymptot. Anal. **32** (1) (2002), 1–26.
- [124] V.Z. Meshkov, *On the possible rate of decrease at infinity of the solutions of second-order partial differential equations*, Mat. Sb. **182** (1991), 364–383 (in Russian); English transl.: Math. USSR-Sb. **72** (1992), 343–361.
- [125] S. Micu and E. Zuazua, *Boundary controllability of a linear hybrid system arising in the control of noise*, SIAM J. Control. Optim. **35** (5) (1997), 1614–1638.
- [126] S. Micu and E. Zuazua, *On the lack of null controllability of the heat equation on the half line*, Trans. Amer. Math. Soc. **353** (4) (2001), 1635–1659.
- [127] S. Micu and E. Zuazua, *An introduction to the controllability of partial differential equations*, Quelques questions de théorie du contrôle, T. Sari, ed., Collection Travaux en Cours, Hermann (2005), 67–150.
- [128] S. Micu and E. Zuazua, *On the controllability of a fractional order parabolic equation*, SIAM J. Control. Optim. **44** (6) (2006), 1950–1972.
- [129] K. Miller, *Nonunique continuation for certain ODE's in Hilbert space and for uniformly parabolic and elliptic equations in self-adjoint divergence form*, Symposium on Non-Well-Posed Problems and Logarithmic Convexity (Heriot-Watt Univ., Edinburgh, 1972), Lecture Notes in Math., Vol. 316, Springer-Verlag, Berlin (1973), 85–101.
- [130] L. Miller, *Escape function conditions for the observation, control, and stabilization of the wave equation*, SIAM J. Control Optim. **41** (5) (2002), 1554–1566.
- [131] L. Miller, *Geometric bounds on the growth rate of null-controllability cost for the heat equation in small time*, J. Differential Equations **204** (1) (2004), 202–226.
- [132] L. Miller, *Controllability cost of conservative systems: Resolvent condition and transmutation*, J. Funct. Anal. **218** (2) (2005), 425–444.
- [133] L. Miller, *On the null-controllability of the heat equation in unbounded domains*, Bull. Sci. Math. **129** (2) (2005), 175–185.
- [134] L. Miller, *On the controllability of anomalous diffusions generated by the fractional Laplacian*, Math. Control Signals Systems **18** (3) (2006), 206–271.
- [135] L. Miller, *On exponential observability estimates for the heat semigroup with explicit rates*, Rend. Lincei Mat. Appl., to appear.
- [136] C. Morawetz, *Notes on time decay and scattering for some hyperbolic problems CBMS-NSF, Regional Conf. Ser. in Appl. Math.*, Vol. 19, SIAM, Philadelphia, PA (1975).
- [137] J.H. Ortega, L. Rosier and T. Takahashi, *Classical solutions for the equations modeling the motion of a ball in a bidimensional incompressible perfect fluid*, M2AN Math. Model. Numer. Anal. **39** (1) (2005), 79108.
- [138] J.H. Ortega, L. Rosier and T. Takahashi, *On the motion of a rigid body immersed in a bidimensional incompressible perfect fluid*, Ann. Inst. H. Poincaré Anal. Non Linéaire, to appear.
- [139] A. Osses, *Une nouvelle famille de multiplicateurs et ses applications à la contrôlabilité exacte des ondes*, C. R. Acad. Sci. Paris **326**, (1998) 1099–1104.
- [140] K.D. Phung, *Observability and control of Schrödinger equations*, SIAM J. Control Optim. **40** (1) (2001), 211–230.
- [141] A. Plis, *On non-uniqueness in Cauchy problems for an elliptic second order differential equation*, Bull. Acad. Polon. Sci. **11** (1963), 95–100.
- [142] J. Ralston, *Solution of the wave equation with localized energy*, Comm. Pure Appl. Math. **22** (1969), 807–823.
- [143] J. Ralston, *Gaussian beams and the propagation of singularities*, Studies in Partial Differential Equations, W. Littman, ed., MAA Stud. in Math., Vol. 23 (1982), 206–248.
- [144] K. Ramdani, T. Takahashi, G. Tenenbaum and M. Tucsnak, *A spectral approach for the exact observability of infinite dimensional systems with skew-adjoint generator*, J. Funct. Anal. **226** (1) (2005), 193–229.
- [145] B.P. Rao and A. Wehbe, *Polynomial energy decay rate and strong stability of Kirchhoff plates with non-compact resolvent*, J. Evol. Equ. **5** (2005), 137–152.
- [146] J. Rauch, X. Zhang and E. Zuazua, *Polynomial decay of a hyperbolic–parabolic coupled system*, J. Math. Pures Appl. **84** (2005), 407–470.
- [147] L. Robbiano, *Fonction de coût et contrôle des solutions des équations hyperboliques*, Asymptot. Anal. **10** (2) (1995), 95–115.

- [148] L. Robbiano and C. Zuily, *Uniqueness in the Cauchy problem for operators with partially holomorphic coefficients*, Invent. Math. **131** (1998), 493–539.
- [149] D.L. Russell, *A unified boundary controllability theory for hyperbolic and parabolic partial differential equations*, Stud. Appl. Math. **52** (1973), 189–221.
- [150] D.L. Russell, *Controllability and stabilizability theory for linear partial differential equations. Recent progress and open questions*, SIAM Rev. **20** (1978), 639–739.
- [151] J. San Martín, V. Starovoitov and M. Tucsnak, *Global weak solutions for the two dimensional motion of several rigid bodies in an incompressible viscous fluid*, Arch. Ration. Mech. Anal. **161** (2) (2002), 113–147.
- [152] T. Seidman, S. Avdonin and S.A. Ivanov, *The “window problem” for series of complex exponentials*, J. Fourier Anal. Appl. **6** (2000), 233–254.
- [153] M. Siegmund, *Null controllability for systems of parabolic equations*, Preprint (2005).
- [154] M. Slemrod, *Weak asymptotic decay via a “relaxed invariance principle” for a wave equation with non-linear, nonmonotone damping*, Proc. Roy. Soc. Edinburgh Sect. A **113** (1989), 87–97.
- [155] E.D. Sontag, *Mathematical Control Theory. Deterministic Finite-Dimensional Systems*, 2nd Edition, Texts Appl. Math., Vol. 6, Springer-Verlag, New York (1998).
- [156] S. Tang and X. Zhang, *Carleman inequality for backward stochastic parabolic equations with general coefficients*, C. R. Math. Acad. Sci. Paris **339** (11) (2004), 775–780.
- [157] D. Tataru, *Unique continuation for solutions to PDE’s: Between Hörmander’s theorem and Holmgren’s theorem*, Comm. Partial Differential Equations **20** (6/7) (1996), 855–884.
- [158] M. Tucsnak, *Contrôle d’une poutre avec actionneur piézoélectrique*, C. R. Acad. Sci. Paris **319** (1994), 697–702.
- [159] M. Tucsnak and G. Weiss, *Passive and conservative linear systems*, Preprint (2006).
- [160] S.R.S. Varadhan, *On the behavior of the fundamental solution of the heat equation with variable coefficients*, Comm. Pure Appl. Math. **20** (1967), 431–455.
- [161] J.L. Vázquez and E. Zuazua, *Lack of collision in a simplified 1-d model for fluid-solid interaction*, Math. Models Methods Appl. Sci. **16** (5) (2006), 637–678.
- [162] P.-F. Yao, *On the observability inequalities for exact controllability of wave equations with variable coefficients*, SIAM J. Control Optim. **37** (5) (1999), 1568–1599.
- [163] R.M. Young, *An Introduction to Nonharmonic Fourier Series*, Academic Press, New York (1980).
- [164] X. Zhang, *Explicit observability estimate for the wave equation with potential and its application*, Roy. Soc. London Proc. Ser. A Math. Phys. Eng. Sci. **456** (2000), 1101–1115.
- [165] X. Zhang and E. Zuazua, *Polynomial decay and control for a 1-d model of fluid-structure interaction*, C. R. Acad. Sci. Paris **336** (2003), 745–750.
- [166] X. Zhang and E. Zuazua, *Control, observation and polynomial decay for a 1-d heat-wave system*, C. R. Acad. Sci. Paris I **336** (2003), 823–828.
- [167] X. Zhang and E. Zuazua, *Polynomial decay and control of a 1-d hyperbolic–parabolic coupled system*, J. Differential Equations **204** (2004), 380–438.
- [168] X. Zhang and E. Zuazua, *Exact controllability of the semilinear wave equation*, Unsolved Problems in Mathematical Systems and Control Theory, Princeton Univ. Press, Princeton, NJ (2004), 173–178.
- [169] X. Zhang and E. Zuazua, *Long time behavior of a coupled heat-wave system arising in fluid–structure interaction*, Preprint (2005); Arch. Ration. Mech. Anal., to appear.
- [170] E. Zuazua, *Exponential decay for semilinear wave equations with localized damping*, Comm. Partial Differential Equations **15** (2) (1990), 205–235.
- [171] E. Zuazua, *Uniform stabilization of the wave equation by nonlinear boundary feedback*, SIAM J. Control Optim. **28** (2) (1990), 466–477.
- [172] E. Zuazua, *Approximate controllability for linear parabolic equations with rapidly oscillating coefficients*, Control Cybernet. **23** (4) (1994), 1–8.
- [173] E. Zuazua, *Exact boundary controllability for the semilinear wave equation*, Nonlinear Partial Differential Equations and Their Applications, Vol. X, H. Brezis and J.-L. Lions, eds, Pitman (1991), 357–391.
- [174] E. Zuazua, *Null controllability of the heat equation in thin domains*, Equations aux dérivées partielles et applications. Articles dédiés à Jacques-Louis Lions, Gauthier-Villars, Paris (1998), 787–801.
- [175] E. Zuazua, *Finite dimensional null controllability for the semilinear heat equation*, J. Math. Pures Appl. **76** (1997), 570–594.

- [176] E. Zuazua, *Some problems and results on the controllability of partial differential equations*, Proceedings of the Second European Conference of Mathematics, Budapest, July 1996, Progr. Math., Vol. 169, Birkhäuser, Basel (1998), 276–311.
- [177] E. Zuazua, *Null control of a 1D model of mixed hyperbolic–parabolic type*, Optimal Control and Partial Differential Equations, J.L. Menaldi et al., eds, IOS Press (2001), 198–210.
- [178] E. Zuazua, *Controllability of partial differential equations and its semidiscrete approximations*, Discrete Contin. Dyn. Syst. **8** (2) (2002), 469–513.
- [179] E. Zuazua, *Optimal and approximate control of finite-difference approximation schemes for the 1D wave equation*, Rend. Mat., Ser. 8 **24** (II) (2004), 201–237.
- [180] E. Zuazua, *Propagation, observation, and control of waves approximated by finite difference methods*, SIAM Rev. **47** (2) (2005), 197–243.
- [181] E. Zuazua, *Control and numerical approximation of the wave and heat*, Proc. Internat. Congress Math., Madrid, 2006, Vol. III, M. Sanz-Solé et al., eds, European Mathematical Society (2006), 1389–1417.

Author Index

Roman numbers refer to pages on which the author (or his/her work) is mentioned. Italic numbers refer to reference pages. Numbers between brackets are the reference numbers. No distinction is made between the first author and co-author(s).

- Abidi, H. 204, 269 [1]
 Agueh, M. 38, 105, 107, 133 [1]; 133 [2]; 133 [3]
 Alarcón, T. 157, 189 [1]; 189 [2]
 Alazard, T. 237, 269 [2]; 269 [3]
 Albert, R. 160, 161, 189 [3]
 Alberti, G. 13, 133 [4]; 290, 380 [1]
 Aleksandrov, A.D. 46, 133 [5]
 Alessandrini, G. 577, 578, 588, 596, 610, 614 [1]
 Alinhac, S. 556, 614 [2]
 Allaire, G. 262, 264, 266, 269 [4]; 269 [5]
 Alt, W. 177, 192 [83]
 Alvarez, O. 509, 524 [2]
 Aly, J.J. 472, 472 [1]; 472 [2]
 Ambrose, D.M. 268, 269 [6]; 269 [7]
 Ambrosi, D. 176, 189 [4]; 190 [37]
 Ambrosio, L. 3, 10, 11, 13, 15, 17, 19, 24, 28, 31, 34, 35, 38, 40, 43, 45, 47, 53, 57, 60, 63, 64, 66, 68, 70, 72, 81, 85, 87, 88, 93, 97, 105, 107, 111, 113, 115, 123, 131, 133, 133 [4]; 133 [6]; 133 [7]; 133 [8]; 133 [9]; 133 [10]; 134 [11]; 281, 290, 308, 310, 311, 317, 319, 321, 323, 349, 350, 352, 354, 357, 358, 365, 378, 380, 380 [2]; 380 [3]; 380 [4]; 380 [5]; 380 [6]; 380 [7]; 381 [8]; 381 [9]; 381 [10]; 381 [11]; 381 [24]; 513, 524 [1]
 Andrzejasson, H. 392, 472 [3]
 Angenent, S. 575, 614 [3]
 Anita, S. 583, 609, 614 [4]; 614 [5]
 Anzellotti, G. 350, 352, 381 [12]; 381 [13]; 381 [14]
 Arnold, A. 107, 134 [12]; 134 [13]
 Arsenov, A.A. 417, 472 [4]
 Asano, K. 248, 275 [167]
 Avdonin, S.A. 552, 569, 583, 607, 608, 614 [6]; 620 [152]
 Babin, A. 238, 269 [8]; 269 [9]; 269 [10]
 Baier, H. 146, 157, 189 [5]
 Baiocchi, C. 106, 134 [14]
 Bally, V. 523, 524 [3]
 Barbu, V. 583, 614 [4]
 Bardos, C. 209, 245, 248, 249, 251, 253, 269 [11]; 269 [12]; 269 [13]; 269 [14]; 269 [15]; 269 [16]; 392, 408, 472 [5]; 546, 547, 561, 604, 614 [7]
 Barkai, N. 167, 189 [6]
 Barkley, D.S. 169, 191 [59]
 Barles, G. 501, 507, 512, 513, 521, 524 [4]; 524 [5]; 524 [6]
 Barlow, M.T. 479, 524 [7]
 Bass, R.F. 479, 524 [7]
 Batt, J. 392, 401, 407, 429, 435, 437, 439, 472 [6]; 473 [7]; 473 [8]; 473 [9]; 473 [10]; 473 [11]; 473 [12]
 Beale, J.T. 238, 270 [23]
 Beauchard, K. 529, 614 [8]; 614 [9]
 Beirão da Veiga, H. 217, 219, 269 [17]; 270 [18]; 270 [19]
 Bellman, R. 479, 524 [8]
 Bellomo, N. 178, 189 [7]; 189 [8]
 Bellouquid, A. 178, 189 [7]; 189 [8]
 Beloussov, L. 170, 190 [41]
 Benameur, J. 244, 270 [20]
 Benamou, J.-D. 5, 30, 134 [15]
 Bžnilan, P. 105, 134 [16]
 Bensoussan, A. 262, 266, 270 [21]; 479, 510, 524 [9]; 524 [10]
 Berestycki, H. 407, 473 [8]
 Berg, H.C. 140, 142, 144, 148, 150, 153, 154, 157, 160, 168, 175, 189 [9]; 189 [10]; 189 [11]; 192 [103]; 192 [106]
 Bernard, P. 18, 134 [17]
 Bers, L. 591, 614 [10]; 614 [11]
 Bertozzi, A.L. 201, 204, 270 [22]; 273 [119]
 Biler, P. 174, 189 [12]; 189 [13]; 189 [14]

- Binney, J. 386, 438, 473 [13]
 Bismut, J.M. 480, 514, 524 [11]; 524 [12]
 Block, S.M. 157, 160, 192 [103]
 Bodart, O. 611, 614 [12]
 Bogachev, V.I. 50, 113, 134 [18]
 Bonhoeffer, F. 146, 157, 189 [5]
 Bonner, J.T. 140, 169, 189 [15]; 189 [16]; 191 [59]
 Bouchard, B. 513, 523, 524 [13]; 524 [14]
 Bouchut, F. 24, 134 [19]; 282, 283, 317, 380 [4];
 381 [15]; 392, 430, 473 [14]; 473 [15]
 Boulakia, M. 530, 597, 613, 614 [13]; 614 [14];
 615 [15]
 Bourgeois, A.J. 238, 270 [23]
 Brakke, K.A. 511, 524 [15]
 Bray, D. 154, 161, 189 [17]; 191 [71]
 Brenier, Y. 5, 17, 30, 134 [15]; 134 [20]; 268,
 270 [24]
 Brenner, P. 335, 381 [16]
 Bresch, D. 237, 244, 256, 270 [25]; 270 [26];
 270 [27]
 Bressan, A. 280, 364, 381 [17]; 381 [18]; 381 [19]
 Brezis, H. 3, 55, 57, 122, 131, 134 [21]; 134 [22];
 134 [23]
 Brose, K. 140, 189 [20]
 Buckdahn, R. 479, 513, 521, 524 [5]; 524 [16];
 524 [17]
 Buffoni, B. 18, 134 [17]
 Burchard, A. 454, 473 [16]
 Burq, N. 546, 548, 561, 615 [16]; 615 [17];
 615 [18]; 615 [19]; 615 [20]; 615 [21]
 Bussolino, F. 176, 189 [4]; 190 [37]
 Buttazzo, G. 45, 134 [24]
- Cabanillas, V. 582, 615 [22]
 Cabre, X. 520, 524 [18]
 Caffarelli, L. 520, 524 [18]
 Caßisch, R.E. 205, 206, 247, 248, 270 [28];
 270 [29]; 274 [150]; 274 [151]
 Caglioti, E. 107, 132, 134 [25]
 Calogero, S. 430, 473 [17]
 Cannarsa, P. 582, 610, 615 [23]; 615 [24]
 Caprette, D.R. 169, 192 [99]
 Cardaliaguet, P. 513, 524 [16]
 Carlen, E.A. 105, 134 [26]; 134 [27]
 Carrillo, J.A. 7, 38, 58, 81, 88, 89, 105, 107, 132,
 134 [28]; 134 [29]; 136 [83]
 Castro, C. 530, 587, 611, 612, 615 [25]; 615 [26];
 615 [27]
 Cercignani, C. 246, 247, 270 [30]; 270 [31]; 386,
 473 [18]
 Chae, D. 574, 615 [28]
 Chalub, F.A. 177, 189 [18]
 Chandrasekhar, S. 156, 189 [19]
- Charvet, N.B. 140, 189 [20]
 Chasseigne, E. 132, 134 [30]
 Chedotal, A. 140, 189 [20]
 Chemin, J.-Y. 201, 204, 207, 238, 244, 270 [32];
 270 [33]; 270 [34]; 270 [35]
 Chen, G.Q. 350, 381 [20]; 381 [21]
 Chen, P. 563, 615 [29]
 Chen, Y.-G. 511D513, 525 [19]
 Cheridito, P. 480, 510, 514D516, 519, 525 [20];
 525 [21]
 Chevance, D. 523, 525 [22]
 Childress, S. 173, 174, 189 [21]; 189 [22]
 Chisholm, R.L. 169, 189 [23]
 Chiu, Y.W. 160, 161, 189 [3]
 Cioranescu, D. 612, 615 [30]; 615 [31]
 Clopeau, T. 209, 270 [36]
 Cohen, M.S. 187, 190 [44]
 Colin, T. 238, 270 [37]
 Coniglio, A. 176, 190 [37]
 Constantin, P. 201, 202, 204, 210, 270 [22];
 270 [38]; 270 [39]; 270 [40]; 270 [41]
 Cordero-Erausquin, D. 107, 134 [31]
 Coron, J.M. 529, 530, 548, 579, 597, 609, 614 [9];
 615 [32]; 615 [33]; 615 [34]; 615 [35];
 615 [36]; 615 [37]; 615 [38]
 Corrias, L. 174, 189 [24]
 Coutand, D. 530, 597, 613, 615 [39]
 Cox, E.C. 181, 188, 191 [67]; 191 [68]; 191 [69];
 192 [88]; 192 [101]
 Cox, S. 587, 616 [40]
 Crandall, M.G. 3, 106, 134 [32]; 134 [33]; 501,
 502, 520, 525 [23]; 525 [24]; 525 [25]
 Crank, J. 148, 189 [25]
 Crippa, G. 281, 283, 284, 349, 350, 352, 354, 357,
 380 [5]; 380 [6]
- Da Prato, G. 113, 134 [34]; 134 [35]
 Dafermos, C. 279, 323, 324, 328, 334, 335,
 381 [22]; 381 [23]
 Dąger, R. 551, 552, 559, 578, 616 [41]
 Dal Maso, G. 290, 380 [7]; 554, 612, 616 [42]
 Dallon, J.C. 181, 189 [26]
 Danchin, R. 204, 215, 217, 223, 269 [1]; 270 [42];
 270 [43]; 270 [44]; 270 [45]; 270 [46]; 271 [47]
 Darcy, H. 264, 265, 271 [48]
 Darroz*, J.-S. 257, 272 [85]
 Davies, E.B. 607, 616 [43]
 de Candia, A. 176, 190 [37]
 de Castro, F. 139, 190 [27]
 De Giorgi, E. 3, 60, 95, 105, 135 [36]; 135 [37];
 283, 381 [24]
 De Lellis, C. 282D284, 290, 308, 323, 365, 380,
 380 [4]; 381 [8]; 381 [9]; 381 [10]; 381 [25];
 381 [26]

- De Lisi, C. 149, 150, 152D156,190 [28]; 190 [29]; 192 [91]
- De Masi, A. 248, 252, 271 [49]
- de Menezes, S. 582, 615 [22]
- de Teresa, L. 582, 611, 616 [45]; 616 [46]; 616 [47]; 616 [48]
- Degiovanni, M. 105, 135 [38]
- Degond, P. 392, 407, 408, 472 [5]; 473 [8]
- Dehman, B. 561, 603, 616 [44]
- Del Grosso, G. 149, 150, 190 [28]
- Del Pino, M. 107, 135 [40]
- Delarue, F. 523, 525 [26]
- Delbrÿck, M. 154,192 [100]
- Delitala, M. 178, 189 [7]
- Dellacherie, C. 10, 135 [39]
- Delort, J.-M. 205, 212, 271 [50]
- Deng, Y. 472, 473 [19]
- Desjardins, B. 221, 224, 228, 229, 231, 237, 244, 245, 270 [25]; 270 [26]; 270 [35]; 271 [51]; 271 [52]; 271 [53]; 271 [54]
- Deutsch, A. 182, 190 [30]
- Devreotes, P.N. 169, 193 [115]
- D'az, J.I. 265, 271 [55]
- Dietz, C. 403, 473 [20]
- DiPerna, R.J. 24, 135 [41]; 212, 245, 251, 254, 271 [56]; 271 [57]; 271 [58]; 271 [59]; 281, 381 [27]; 386, 392, 473 [21]; 473 [22]
- DiTalia, S. 176, 190 [37]
- Dodd, J. 140, 193 [113]
- Dolbeault, J. 107, 134 [12]; 135 [40]; 135 [42]; 136 [83]; 393, 431, 433, 473 [23]; 473 [24]
- Donato, P. 612, 615 [30]
- Dormann, S. 182, 190 [30]
- Dormy, E. 244, 271 [51]
- Doubova, A. 577, 583, 587, 596, 597, 616 [49]; 616 [50]; 616 [51]
- Douglas, J. 514, 525 [27]
- Doyle, J. 167, 168, 193 [120]
- Dunbar, S.R. 177, 192 [83]
- Duyckaerts, Th. 579, 583D585, 606,616 [52]; 616 [53]
- Dynkin, E.B. 479, 525 [28]; 525 [29]; 525 [30]; 525 [31]
- E, W. 205, 206, 271 [60]; 271 [61]
- Ebin, D.G. 215, 217, 271 [62]
- Ehlers, J. 407, 473 [25]
- Eisenbach, M. 140, 190 [31]
- Ekeland, Y. 575, 616 [54]
- El Karoui, N. 514, 525 [32]
- Embid, P.F. 238, 271 [63]
- Engquist, B. 206, 271 [61]
- Erban, R. 178, 190 [32]; 190 [33]
- Escalante, R. 140, 169, 190 [34]
- Escauriaza, L. 573, 577, 578, 588, 596, 610, 614 [1]; 616 [55]
- Esposito, R. 248, 252, 271 [49]
- Evans, L.C. 12, 13, 105, 135 [43]; 135 [44]; 501, 511D513,525 [23]; 525 [33]
- Fabre, C. 530, 564, 575, 582, 616 [56]; 616 [57]; 616 [58]
- Fabrie, P. 238, 270 [37]
- Faltenbacher, W. 437, 473 [9]
- Fattorini, H. 530, 568, 573, 616 [59]; 616 [60]; 616 [61]
- Federer, H. 11, 12, 135 [45]
- Feireisl, E. 201, 223, 224, 233, 271 [64]; 271 [65]; 271 [66]; 271 [67]; 562, 616 [62]
- Feng, J. 105, 135 [46]
- Fernández-Cara, E. 530, 564, 572, 576D579, 581D584, 587, 588, 593, 595D597, 607, 609D611,616 [49]; 616 [50]; 616 [63]; 616 [64]; 617 [65]; 617 [66]; 617 [67]; 617 [68]; 617 [69]; 617 [70]; 617 [71]
- Feynman, R.P. 479, 481, 525 [34]
- Fiedler, B. 184, 190 [35]
- Firtel, R.A. 169, 189 [23]
- Fleming, W.H. 479, 480, 496, 498D503, 507, 509, 510, 520, 525 [35]; 525 [36]
- Foias, C. 201, 210, 270 [39]
- Freidlin, M.I. 479, 490, 525 [37]; 525 [38]
- Freistÿhler, H. 279,381 [28]; 381 [29]; 381 [30]; 381 [31]
- Frid, H. 323, 324, 326, 350, 381 [20]; 381 [21]; 381 [32]
- Fridman, A.M. 386, 438, 473 [26]
- Friedlander, S.K. 156, 190 [36]
- Fu, X. 578, 617 [72]
- Fujiwara, T. 491, 525 [39]
- Fursikov, A.V. 529, 569, 572, 574, 578D580, 594, 597, 615 [36]; 617 [73]; 617 [74]; 617 [75]; 617 [76]; 617 [77]
- Fusco, N. 28, 43, 53, 72, 133 [8]; 283D290, 311, 352, 358, 378, 381 [11]
- Gagliardo, E. 358, 381 [33]
- Galko, M.J. 139, 192 [104]
- Gallagher, I. 215, 222, 238, 244, 270 [35]; 271 [68]; 271 [69]; 271 [70]
- Gamba, A. 176, 177, 189 [4]; 190 [37]; 191 [61]
- Gangbo, W. 14, 17, 105, 134 [26]; 134 [27]; 135 [43]; 135 [47]; 135 [48]
- Garcia, G.C. 611, 616 [63]
- Gardiner, C.W. 145, 179, 190 [38]
- Gardner, R. 38, 135 [49]
- Gariepy, R.F. 12, 13, 135 [44]

- Garofalo, N. 228, 271 [71]
 Gasser, I. 430, 473 [27]
 Gentil, I. 107, 135 [40]
 GŹrard, P. 237, 271 [72]; 546, 615 [19]; 617 [78]
 GŹrard-Varet, D. 238, 244, 270 [25]; 271 [73]
 Ghidaglia, J.-M. 564, 617 [79]
 Ghoussoub, N. 107, 133 [3]
 Giacomelli, L. 105, 135 [50]; 135 [51]
 Gianazza, U. 105D107, 135 [52]
 Giaquinta, M. 69, 135 [53]
 Gierer, A. 161, 190 [39]
 Giga, Y. 511D513, 525 [19]
 Gigli, N. 3D7, 11, 14, 15, 17D19, 34, 35, 38, 40, 45D47, 57, 60, 63, 64, 66, 68, 70, 81, 85, 87, 88, 93, 97, 105D107, 111, 113, 115, 123, 131D133, 133 [9]
 Giraudo, E. 176, 189 [4]; 190 [37]
 Glasner, K. 105, 135 [54]
 Glass, O. 530, 597, 617 [80]
 Glassey, R.T. 386, 392, 401, 407, 413, 431, 433, 473 [28]; 473 [29]; 473 [30]; 473 [31]; 473 [32]; 473 [33]; 473 [34]; 473 [35]; 474 [36]; 474 [37]
 Goffman, C. 45, 135 [55]
 Goldbeter, A. 181, 191 [66]; 191 [74]
 Goldstein, B. 155, 192 [93]
 Goldstein, R.E. 181, 191 [67]; 191 [68]; 191 [69]
 Golse, F. 24, 134 [19]; 245, 248, 249, 251D254, 259, 269 [12]; 269 [13]; 269 [14]; 269 [15]; 271 [74]; 272 [75]; 272 [76]; 272 [77]; 272 [78]; 272 [79]; 392, 473 [15]
 Gomer, R.H. 169, 192 [99]
 Gonz lez-Burgos, M. 572, 583, 595, 611, 614 [12]; 616 [50]; 616 [64]; 617 [81]
 Goodhill, G.J. 153, 154, 190 [40]
 Goodman, C.S. 140, 189 [20]
 Gordon, R. 170, 190 [41]
 Goto, S. 511D513, 525 [19]
 Greenspan, H. 201, 238, 272 [80]
 Grenier, E. 205, 206, 221, 224, 228, 229, 231, 237, 238, 244, 245, 270 [26]; 270 [35]; 271 [51]; 271 [52]; 271 [53]; 271 [54]; 272 [81]; 272 [82]; 272 [83]; 272 [84]
 Grisvard, P. 548, 617 [82]
 Grotta Ragazzo, C. 437, 474 [45]
 Guerrero, S. 530, 548, 572, 579, 595D597, 610, 611, 615 [37]; 616 [64]; 617 [65]; 617 [66]; 617 [67]; 617 [68]; 617 [83]
 Gugat, M. 557, 558, 608, 609, 617 [84]; 617 [85]
 Guill n-Gonz lez, F. 256, 270 [27]
 Guiraud, J.-P. 257, 272 [85]
 Guo, Y. 254, 272 [86]; 386, 430, 435, 437, 442, 454, 464, 469, 470, 472, 473 [16]; 474 [38]; 474 [39]; 474 [40]; 474 [41]; 474 [42]; 474 [43]; 474 [44]; 474 [45]; 474 [46]; 474 [47]; 474 [48]; 474 [49]; 474 [50]; 474 [51]; 474 [52]
 Had i , M. 454, 474 [53]
 Hagan, P.S. 183, 185, 187, 190 [42]; 190 [43]; 190 [44]
 Hagstrom, T. 217, 221, 222, 272 [87]
 Halloy, J. 181, 191 [66]
 Hansen, S. 559, 562, 617 [86]
 Haraux, A. 552, 559, 560, 562, 617 [87]; 617 [88]; 617 [89]
 Hauray, M. 387, 474 [54]
 Heinrich, R. 157, 158, 160, 190 [45]
 Heinzle, M. 437, 474 [55]
 Herrero, M.A. 140, 169, 170, 174, 178, 180, 181, 184, 188, 189 [8]; 190 [46]; 190 [47]; 190 [48]; 190 [49]
 Hesla, T.I. 597, 617 [90]
 Hilbert, D. 245, 272 [88]
 Hildebrandt, S. 69, 135 [53]
 Hillairet, M. 597, 617 [91]
 Hillen, T. 175D178, 190 [50]; 192 [84]; 192 [87]
 Ho, L.F. 544, 617 [92]
 Hoff, D. 217, 221, 223, 272 [89]
 Holm, D.D. 437, 440, 474 [56]
 Hopf, E. 210, 272 [90]
 H rmander, L. 554, 617 [93]
 Horst, E. 393, 407, 415, 417, 429, 435, 437, 473 [9]; 474 [57]; 474 [58]; 474 [59]; 474 [60]; 474 [61]
 Horstmann, D. 170, 174, 190 [51]
 Howard, L.N. 182, 190 [52]
 Huang, C. 105, 135 [56]
 Huang, Y. 167, 168, 193 [120]
 Hunze, R. 417, 474 [61]
 Hwang, H.J. 430, 474 [62]
 Ibrahim, S. 244, 270 [20]
 Iftimie, D. 268, 272 [91]
 Iglesias, P.A. 164D166, 190 [53]; 191 [62]; 191 [70]
 Ikeda, N. 518, 525 [40]
 Illner, R. 246, 247, 270 [31]; 386, 393, 431, 473 [18]; 474 [63]
 Imanuvilov, O.Yu. 529, 530, 548, 569, 572, 574, 578D580, 583, 588, 589, 594, 595, 597, 615 [28]; 617 [68]; 617 [75]; 617 [76]; 617 [77]; 618 [94]; 618 [95]; 618 [96]; 618 [97]; 618 [98]
 Ingham, A.E. 550, 618 [99]
 Isakov, V. 550, 618 [100]
 Ishii, H. 502, 520, 525 [24]; 525 [41]
 Isozaki, H. 220, 272 [92]; 272 [93]

- Ivanov, S.A. 552, 569, 583, 607, 608, 614 [6]; 620 [152]
- Jabin, P.-E. 387, 430, 473 [27]; 474 [54]
- Jaffard, S. 552, 617 [88]
- JŠger, W. 174, 190 [54]
- Jang, J. 472, 474 [64]
- Jang, W. 169, 192 [99]
- Jeans, J. 386, 474 [65]
- Jessell, T.M. 139, 140, 192 [104]; 193 [113]
- John, F. 147, 191 [55]; 591, 614 [10]
- Joly, J.-L. 220, 272 [94]
- Jordan, R. 4, 30, 99, 105, 109, 113, 135 [56]; 135 [57]
- Jost, J. 46, 135 [58]
- Kac, M. 479, 481, 525 [42]
- Kang, X. 107, 133 [3]
- Karatzas, I. 481, 483, 486, 499, 525 [43]
- Karch, G. 174, 189 [13]; 189 [14]
- Kato, T. 202, 207, 272 [95]; 272 [96]
- Katsoulakis, M. 105, 135 [46]
- Kavian, O. 611, 616 [47]
- Keller, E.F. 170, 171, 175, 191 [56]; 191 [57]
- Kennedy, T.E. 139, 192 [104]
- KeyPtz, B.L. 279, 381 [34]; 382 [35]
- Kidd, T. 140, 189 [20]
- Kim, K.K. 154, 160, 161, 191 [58]
- Kim, S.H. 154, 160, 161, 191 [58]
- Kim, S.M. 574, 615 [28]
- Kinderlehrer, D. 4, 30, 99, 105, 107, 109, 113, 135 [42]; 135 [57]; 135 [59]
- Klainerman, S. 215, 217, 219, 220, 272 [97]; 272 [98]; 392, 475 [66]
- Knott, M. 17, 135 [60]
- Koga, S. 182, 191 [64]
- Komornik, V. 545, 551, 552, 560, 618 [101]; 618 [102]; 618 [103]
- Komura, Y. 3, 135 [61]
- Konijn, T.M. 169, 191 [59]; 191 [75]
- Koppel, N. 182, 190 [52]
- Kowalczyk, M. 107, 135 [42]
- Kowalczyk, R. 177, 191 [60]; 191 [61]
- Krabs, W. 609, 618 [104]
- Kranzer, H.C. 279, 381 [34]
- Kreiss, H.-O. 217, 272 [99]
- Krishnan, J. 164, 165, 191 [62]
- Kruse, K.-O. 469, 475 [67]
- Kruzhkov, S. 322, 382 [36]
- Krylov, N.V. 479, 520, 525 [44]; 525 [45]
- Kunii, H. 175, 193 [110]
- Kunita, H. 491, 525 [39]
- Kunze, M. 430, 435, 473 [10]; 475 [68]
- Kuramoto, Y. 182, 191 [63]; 191 [64]
- Kurth, R. 392, 405, 475 [69]; 475 [70]
- Lachowicz, M. 248, 272 [100]
- Ladyzenskaya, O.A. 520, 525 [46]
- Landau, L.D. 391, 475 [71]
- Lanford, O.E., III 246, 272 [101]
- Lasiecka, I. 530, 618 [105]
- Lauffenburger, D.A. 143Ð145, 149, 191 [65]
- Lauren•ot, Ph. 174, 189 [13]; 189 [14]
- Lauzeral, J. 181, 191 [66]
- Lebeau, G. 542, 546, 547, 553, 561, 562, 569, 570, 572, 598, 603, 604, 608, 610, 612, 614 [7]; 615 [20]; 616 [44]; 618 [106]; 618 [107]; 618 [108]; 618 [109]; 618 [110]
- Lebowitz, J.L. 248, 252, 271 [49]
- Lee, E.B. 529, 531, 618 [111]
- Lee, K.J. 181, 191 [67]; 191 [68]; 191 [69]
- Leibler, S. 167, 189 [6]
- Lemou, M. 435, 467, 469, 475 [72]
- Leray, J. 210, 215, 273 [102]; 273 [103]; 273 [104]
- Leugering, G. 557, 558, 608, 609, 617 [84]; 617 [85]
- Levchenko, A. 164, 166, 190 [53]; 191 [70]
- Levermore, D. 245, 248, 249, 251Ð254, 259, 269 [12]; 269 [13]; 269 [14]; 269 [15]; 271 [74]; 273 [105]
- Levin, M.D. 154, 161, 189 [17]; 191 [71]
- Levine, H. 165Ð167, 192 [98]
- Liboff, R.L. 386, 475 [73]
- Lieb, E.H. 390, 426, 442, 450, 453, 457, 468, 475 [74]
- Lifschitz, E.M. 391, 475 [71]
- Liggett, T.M. 106, 134 [32]
- Lin, C.K. 217, 237, 270 [26]; 273 [106]
- Lin, Z. 469, 475 [75]; 475 [76]; 475 [77]; 475 [78]
- Linderman, J.J. 143Ð145, 149, 191 [65]
- Lions, J.-L. 262, 266, 270 [21]; 273 [107]; 479, 510, 524 [9]; 524 [10]; 529, 542, 543, 545, 554, 555, 560, 564, 565, 611, 618 [112]; 618 [113]; 618 [114]; 618 [115]; 618 [116]
- Lions, P.-L. 24, 135 [41]; 201, 214, 215, 217, 223Ð225, 227, 228, 230, 233, 236, 245, 248, 249, 251, 252, 254, 260, 265, 267, 271 [54]; 271 [56]; 272 [75]; 273 [108]; 273 [109]; 273 [110]; 273 [111]; 273 [112]; 273 [113]; 273 [114]; 273 [115]; 273 [116]; 281, 381 [27]; 386, 392, 417, 423, 460, 473 [21]; 473 [22]; 475 [79]; 475 [80]; 479, 493, 494, 501, 502, 520, 523, 525 [23]; 525 [24]; 525 [25]; 525 [47]; 526 [48]; 526 [49]; 526 [50]; 526 [51]
- Lisini, S. 105, 133 [10]
- Liu, K. 548, 560, 618 [117]; 618 [118]
- Liu, T.-P. 472, 473 [19]

- Liu, W.J. 559, 618 [119]
 Liu, Z. 548, 600, 618 [118]; 618 [120]
 Loeper, G. 402, 475 [81]
 Loomis, W.F. 165Ð167, 192 [98]
 López, A. 548, 611, 612Ð18 [121]; 618 [122]
 Lorenz, J. 217, 221, 222, 272 [87]
 Loreti, P. 551, 552, 618 [102]
 Losada, A. 169, 191 [75]
 Loss, M. 390, 426, 442, 450, 453, 457, 468, 475 [74]
 Lott, J. 107, 135 [62]
 Luckhaus, S. 174, 175, 190 [54]; 191 [72]
 Lumsden, A.G. 140, 193 [113]
 Lunardi, A. 113, 134 [34]

 Ma, J. 479, 514, 524 [17]; 525 [27]; 526 [52]; 526 [53]
 Ma, L. 164, 165, 191 [62]
 Ma, Z.-M. 123, 135 [63]
 Maciá, F. 547, 577, 587, 603, 619 [123]
 Maeda, M. 140, 191 [73]
 Mahalov, A. 238, 269 [8]; 269 [9]; 269 [10]
 Majda, A.J. 201, 212, 215, 217, 219, 220, 238, 271 [57]; 271 [58]; 271 [59]; 271 [63]; 272 [97]; 272 [98]; 273 [117]; 273 [118]; 273 [119]; 402, 475 [82]; 475 [83]
 Majda, G. 402, 475 [82]
 Majdoub, M. 244, 270 [20]
 Malgrange, B. 564, 618 [116]
 Mal, J. 283, 284, 308, 365, 381 [10]
 Maniglia, S. 283, 284, 349, 350, 352, 354, 357, 380 [6]
 Marchetti, F. 149, 150, 152, 153, 190 [28]; 190 [29]
 Marchioro, C. 201, 268, 273 [120]; 273 [121]; 273 [122]
 Marillat, V. 140, 189 [20]
 Marino, A. 3, 60, 105, 135 [37]; 135 [38]; 135 [64]
 Markowich, P.A. 107, 134 [13]; 136 [83]; 177, 189 [18]
 Markus, L. 529, 531, 618 [111]
 Marsden, J.E. 437, 440, 474 [56]
 Martiel, J.L. 181, 191 [74]
 Martinez, P. 582, 610, 615 [23]; 615 [24]
 Masmoudi, N. 204, 209, 210, 215, 217, 224, 227Ð229, 232, 233, 236, 238, 241Ð245, 248, 249, 251Ð254, 256, 258, 260, 267, 268, 269 [7]; 270 [27]; 271 [54]; 272 [84]; 273 [105]; 273 [111]; 273 [112]; 273 [113]; 273 [114]; 273 [115]; 273 [116]; 273 [123]; 273 [124]; 273 [125]; 273 [126]; 273 [127]; 273 [128]; 274 [129]; 274 [130]; 274 [131]; 274 [132]
 Mato, J.M. 169, 191 [75]
 Mawhin, J. 446, 475 [84]
 Mayer, U.F. 106, 136 [65]
 McCann, R.J. 6, 7, 17, 38, 43, 58, 81, 88, 89, 105, 107, 132, 134 [28]; 134 [29]; 135 [48]; 136 [66]; 136 [67]
 Medina, E. 174, 190 [47]
 Mĭzhats, F. 435, 467, 469, 475 [72]
 Meinhardt, H. 161Ð164, 190 [39]; 191 [76]; 191 [77]; 191 [78]
 Menozzi, F. 523, 525 [26]
 Meshkov, V.Z. 585, 619 [124]
 Mĭžtĭvier, G. 217, 220, 236, 237, 272 [94]; 274 [133]; 274 [134]
 Meyer, P.-A. 10, 135 [39]
 Micu, S. 530, 551, 552, 554, 572, 573, 582, 619 [125]; 619 [126]; 619 [127]; 619 [128]
 Mikami, T. 105, 136 [68]
 Mikeliū, A. 209, 262, 270 [36]; 274 [135]
 Mikhailov, A.S. 182, 191 [79]
 Miller, K. 591, 619 [129]
 Miller, L. 545, 547, 548, 572, 573, 578, 582, 583, 607, 608, 619 [130]; 619 [131]; 619 [132]; 619 [133]; 619 [134]; 619 [135]
 Mirzayan, C. 139, 192 [104]
 Mischler, S. 257, 274 [136]
 Mora, C.A. 279, 382 [35]
 Morawetz, C. 544, 619 [136]
 Morrison, P. 437, 439, 473 [11]
 Morton-Firth, C.J. 161, 189 [17]
 Murat, F. 612, 615 [31]

 Nadzieja, T. 174, 189 [13]; 189 [14]
 Nagai, T. 174, 191 [80]; 191 [81]
 Nakanishi, K. 268, 274 [130]; 274 [131]
 Nanjundiah, V. 140, 169, 172, 191 [75]; 191 [82]
 Nazaret, B. 107, 134 [31]
 Neel, B.G. 157, 158, 160, 190 [45]
 Neunzert, H. 387, 475 [85]
 Nguetseng, G. 266, 274 [137]; 274 [138]
 Nicolaenko, B. 238, 269 [8]; 269 [9]; 269 [10]
 Nirenberg, L. 591, 614 [11]
 Nochetto, R.H. 106, 136 [69]
 Novotnĕ, A. 201, 223, 274 [139]

 Oksendal, B. 491, 492, 509, 510, 526 [54]
 Oleinik, O.A. 206, 274 [140]; 274 [141]; 274 [142]
 Ortega, J.H. 597, 619 [137]; 619 [138]
 Osses, A. 530, 545, 577, 587, 596, 597, 611, 615 [15]; 616 [51]; 616 [63]; 619 [139]
 Othmer, H.G. 160, 161, 175, 177, 178, 181, 189 [3]; 189 [26]; 190 [32]; 190 [33]; 190 [50]; 192 [83]; 192 [84]; 192 [85]; 192 [86]; 192 [107]; 193 [112]

- Otto, F. 4, 7, 30, 38, 46, 88, 99, 105, 107, 109, 113, 131, 135 [50]; 135 [51]; 135 [57]; 136 [70]; 136 [71]; 136 [72]; 136 [73]; 136 [74]; 136 [75]; 136 [76]
- Page, K.M. 157, 189 [1]; 189 [2]
- Pag•s, G. 523, 524 [3]
- Paicu, M. 238, 274 [143]
- Painter, K.J. 175, 176, 192 [87]
- Pallara, D. 28, 43, 53, 72, 133 [8]; 283D290, 311, 352, 358, 378, 381 [11]
- Pallard, C. 392, 430, 473 [15]; 475 [86]
- Palsson, E. 181, 192 [88]
- Papanicolaou, G. 262, 266, 270 [21]
- Pardoux, E. 514, 521, 522, 524 [5]; 526 [55]; 526 [56]; 526 [57]
- Parkinson, J.S. 178, 192 [107]
- Pazy, A. 3, 128, 134 [33]; 136 [77]
- Pedlovsky, J. 201, 209, 238, 243, 274 [144]
- Peng, S. 480, 514, 522, 525 [32]; 526 [55]; 526 [56]; 526 [57]; 526 [58]; 526 [59]
- Percus, J.K. 173, 189 [22]
- Perelson, A.S. 154D156, 192 [89]; 192 [90]; 192 [91]
- Přez-Garcia, R. 611, 614 [12]; 617 [81]
- Perthame, B. 174, 177, 189 [18]; 189 [24]; 192 [92]; 251, 272 [75]; 272 [76]; 392, 393, 407, 417, 423, 430, 431, 473 [8]; 473 [27]; 475 [80]; 475 [87]; 475 [88]
- Petzeltov†, H. 233, 271 [67]
- Pfaffelmoser, K. 392, 417, 475 [89]
- Phung, K.D. 572, 578, 619 [140]
- Pierre, M. 132, 136 [78]
- Placzek, M. 140, 193 [113]
- Plis, A. 591, 619 [141]
- Polyachenko, V.L. 386, 438, 473 [26]
- Posner, R.G. 155, 192 [93]
- Prandtl, L. 205, 274 [145]
- Pratelli, A. 14, 136 [79]
- Preziosi, L. 176, 177, 189 [4]; 190 [37]; 191 [61]
- Protter, P. 514, 525 [27]; 526 [52]; 526 [53]
- Puel, J.-P. 530, 564, 572, 575, 577, 579, 582, 587, 588, 595D597, 616 [51]; 616 [56]; 616 [57]; 616 [58]; 616 [64]; 617 [68]; 618 [96]; 618 [97]
- Pulvirenti, M. 24, 134 [19]; 201, 246, 247, 268, 270 [31]; 273 [121]; 273 [122]; 386, 473 [18]
- Purcell, E.M. 142, 144, 148D150, 153, 154, 157, 189 [11]
- Ralston, J. 546, 603, 619 [142]; 619 [143]
- Ramdani, K. 548, 619 [144]
- Ram—n y Cajal, S. 139192 [94]; 192 [95]; 192 [96]
- Rao, B.P. 548, 562, 600, 618 [118]; 618 [120]; 619 [145]
- Raper, K.B. 140, 192 [97]
- Raphael, P. 435, 467, 469, 475 [72]
- Rapoport, T.A. 157, 158, 160, 190 [45]
- Rappel, W.J. 165D167, 192 [98]
- Ratiu, T. 437, 440, 474 [56]
- Rauch, J. 220, 272 [94]; 335, 382 [37]; 546, 547, 561, 598, 601, 603, 604, 613, 614 [7]; 619 [146]
- Raugel, G. 268, 272 [91]; 274 [146]
- Regnier, H. 523, 526 [48]
- Rein, G. 392, 393, 401, 413, 429, 431, 434D437, 439, 440, 442, 443, 454, 460, 467, 469, 470, 472, 473 [10]; 473 [11]; 473 [12]; 473 [23]; 474 [46]; 474 [47]; 474 [48]; 474 [49]; 474 [50]; 474 [63]; 475 [67]; 475 [90]; 475 [91]; 475 [92]; 475 [93]; 476 [94]; 476 [95]; 476 [96]; 476 [97]; 476 [98]; 476 [99]; 476 [100]; 476 [101]; 476 [102]
- Rendall, A.D. 392, 401, 413, 429, 430, 436, 437, 469, 474 [55]; 475 [68]; 476 [100]; 476 [101]; 476 [102]
- Rienstra, W. 407, 473 [25]
- Robbiano, L. 542, 553, 556, 569, 570, 572, 608, 610, 613, 618 [108]; 619 [147]; 620 [148]
- Robert, R. 209, 270 [36]; 402, 476 [103]
- Rockafellar, R.T. 57, 136 [80]
- Ršckner, M. 123, 135 [63]
- Rodr'guez-Bellido, M.A. 256, 270 [27]
- Roisin-Bouffey, C. 169, 192 [99]
- Rosier, L. 597, 619 [137]; 619 [138]
- Rossi, R. 106, 136 [81]
- Rousset, F. 245, 274 [147]; 274 [148]
- Rulla, J. 106, 136 [82]
- Russell, D.L. 529, 552, 560, 568, 569, 572, 573, 616 [60]; 616 [61]; 620 [149]; 620 [150]
- Saccon, C. 3, 105, 135 [64]
- Saffman, P.G. 154, 192 [100]
- Saint-Raymond, L. 245, 248, 249, 251D254, 258, 260, 272 [77]; 272 [78]; 272 [79]; 274 [132]; 274 [149]
- Sammartino, M. 205, 206, 270 [29]; 274 [150]; 274 [151]
- Samokhin, V.N. 206, 274 [142]
- San Martin, J. 514, 526 [52]; 530, 620 [151]
- S•nchez (Sanchez), O. 393, 433, 467, 473 [24]; 476 [104]
- S•nchez-Palencia, E. 262, 266, 274 [152]
- Sandor, V. 403, 473 [20]
- Qin, H. 563, 615 [29]
- Quenez, M.C. 514, 525 [32]
- Quimcampoix, M. 513, 524 [16]

- Sastre, L. 140, 169, 170, 180, 181, 188, 190 [48]
 Savarž, G. 3Đ7, 11, 14, 15, 17Đ19, 34, 35, 38, 40, 45Đ47, 57, 60, 63, 64, 66, 68, 70, 81, 85, 87, 88, 93, 97, 105Đ107, 111, 113, 115, 123, 131Đ133, 133 [9]; 133 [10]; 135 [52]; 136 [69]; 136 [81]
 Savin, O. 105, 135 [43]
 Sawal, S. 181, 188, 192 [101]
 Sayah, A. 510, 526 [60]
 Schaap, P. 181, 192 [85]
 Schaeffer, J. 392, 407, 413, 418, 433, 465, 467, 473 [29]; 473 [30]; 473 [31]; 473 [32]; 473 [33]; 473 [34]; 476 [105]; 476 [106]; 476 [107]
 Schechter, M. 591, 614 [10]
 Scheel, A. 184, 190 [35]; 192 [102]
 Schmeiser, C. 177, 189 [18]
 Schochet, S. 215, 217, 220, 236Đ238, 274 [133]; 274 [134]; 274 [153]; 274 [154]; 274 [155]
 Seg^la, F. 228, 271 [71]
 Segall, J.E. 157, 160, 192 [103]
 Segel, L.A. 170, 171, 175, 191 [57]
 Seidman, T. 583, 607, 608, 620 [152]
 Sell, G.R. 268, 274 [146]
 Senba, T. 174, 191 [81]
 Sentis, R. 251, 272 [75]; 272 [76]
 Seraċni, T. 139, 192 [104]
 Seregin, G. 573, 616 [55]
 Serini, G. 176, 189 [4]; 190 [37]
 Serre, D. 279, 328, 382 [38]
 Serrin, J. 45, 135 [55]
 Shimizu, T.S. 154, 191 [71]
 Shkoller, S. 530, 597, 613, 615 [39]
 Shreve, S. 481, 483, 486, 499, 525 [43]
 Siegwart, M. 611, 620 [153]
 Simon, M.L. 167, 168, 193 [120]
 Sklyar, G. 557, 608, 617 [85]
 Skorokhod, A.V. 491, 492, 526 [61]
 Slemrod, M. 602, 620 [154]
 Smith, C.S. 17, 135 [60]
 Soler, J. 393, 433, 467, 473 [24]; 476 [104]
 Solonnikov, V.A. 520, 525 [46]
 Sone, Y. 246, 274 [156]
 Soner, H.M. 479, 480, 496Đ503, 507, 509Đ516, 519, 520, 524 [1]; 524 [6]; 525 [20]; 525 [21]; 525 [35]; 526 [62]; 526 [63]; 526 [64]; 526 [65]; 526 [66]; 526 [67]
 Sontag, E.D. 531, 620 [155]
 Sotelo, C. 139, 140, 189 [20]; 192 [105]
 Souganidis, P.E. 479, 480, 512, 513, 524 [6]; 525 [36]; 526 [49]; 526 [50]
 Sourjik, V. 160, 192 [106]
 Sparber, C. 107, 136 [83]
 Spiro, P. 178, 192 [107]
 Spohn, H. 246, 275 [157]
 Spruck, L. 511Đ513, 525 [33]
 Stafflani, G. 392, 475 [66]
 Starovoitov, V. 530, 620 [151]
 Stein, E. 424, 425, 476 [108]
 Stevens, A. 175, 178Đ180, 192 [86]; 192 [108]
 Straskraba, I. 201, 223, 274 [139]
 Strauss, W.A. 392, 401, 413, 431, 469, 473 [35]; 474 [36]; 474 [37]; 474 [51]; 474 [52]; 475 [77]; 475 [78]
 Strichartz, R.S. 221, 275 [158]
 Sturm, K. 107, 136 [84]
 Sugiyama, Y. 175, 191 [72]; 192 [109]; 193 [110]
 Sulem, A. 491, 492, 509, 510, 526 [54]
 Suzuki, T. 170, 174, 191 [81]; 193 [111]
 Sverċk, V. 573, 616 [55]
 Swann, H.S.G. 202, 275 [159]
 Sznitman, A.S. 493, 526 [51]
 Takahashi, T. 548, 597, 619 [137]; 619 [138]; 619 [144]
 Talay, D. 514, 522, 526 [68]
 Tang, S. 613, 620 [156]
 Tang, Y. 181, 193 [112]
 Tartar, L. 237, 262Đ265, 275 [160]; 275 [161]
 Tataru, D. 556, 609, 614 [5]; 620 [157]
 Temam, R. 201, 207Đ210, 275 [162]; 275 [163]; 275 [164]; 275 [165]; 575, 616 [54]
 Tenenbaum, G. 548, 619 [144]
 Tessier-Lavigne, M. 139, 140, 189 [20]; 192 [104]; 193 [113]
 Thomas, P.J. 165Đ167, 192 [98]
 Thomson, P.A. 181, 188, 192 [101]
 Tilli, P. 14, 31, 134 [11]
 Torres, S. 514, 526 [52]
 Toscani, G. 105Đ107, 134 [13]; 135 [52]
 Tosques, M. 3, 60, 105, 135 [37]; 135 [38]; 135 [64]
 Tourin, A. 509, 524 [2]
 Touzi, N. 480, 497, 499, 500, 503, 510Đ512, 514Đ516, 519, 523, 524 [14]; 525 [20]; 525 [21]; 526 [64]; 526 [65]; 526 [66]; 526 [67]
 Trċlat, E. 609, 615 [38]
 Tremaine, S. 386, 438, 473 [13]
 Triggiani, R. 530, 618 [105]
 Tucsnak, M. 530, 548, 549, 552, 619 [144]; 620 [151]; 620 [158]; 620 [159]
 Turing, A.M. 162, 170, 193 [114]
 Uggl, C. 437, 474 [55]
 Ukai, S. 217, 220, 231, 248, 269 [16]; 275 [166]; 275 [167]
 Unterreiter, A. 107, 134 [13]
 Uraltseva, N.N. 520, 525 [46]
 Urbach, J.S. 153, 154, 190 [40]

- van de Meere, J.G.C. 169, 191 [59]
 van Haastert, P.J.M. 169, 193 [115]
 van Ooijen, A. 168, 179, 193 [116]
 Vancostenoble, J. 582, 610, 615 [23]; 615 [24]
 Varadhan, S.R.S. 607, 620 [160]
 Vasseur, A. 322, 382 [39]
 Vázquez (Vazquez), J.L. 132, 134 [30]; 597, 620 [161]
 Velázquez, J.J.L. 174, 175, 190 [47]; 190 [49]; 193 [117]; 193 [118]
 Verdi, C. 106, 136 [69]
 Vicente, J.J. 140, 169, 190 [34]
 Victor, N. 480, 514Ð516, 525 [21]
 Villani, C. 7, 15, 18, 38, 50, 58, 81, 88, 89, 105, 107, 132, 134 [25]; 134 [28]; 134 [29]; 134 [31]; 135 [62]; 136 [75]; 136 [85]; 136 [86]; 246, 247, 275 [168]
 Vlasov, A.A. 386, 476 [109]; 476 [110]

 Walkington, N.J. 107, 135 [59]
 Wang, K. 140, 189 [20]
 Wang, W. 154, 160, 161, 191 [58]
 Wang, X. 208, 209, 268, 275 [164]; 275 [165]; 275 [169]; 275 [170]
 Watanabe, S. 518, 525 [40]
 Weckler, J. 430, 476 [111]
 Wehbe, A. 562, 619 [145]
 Weinstein, A. 437, 440, 474 [56]
 Weiss, G. 549, 620 [159]
 Wentzell, A.D. 479, 525 [38]
 Westdickenberg, M. 7, 88, 136 [76]
 Wets, R.J.-B. 57, 136 [80]
 Willem, M. 446, 475 [84]
 Wofsy, C. 155, 192 [93]
 Wolansky, G. 444, 469, 476 [112]; 476 [113]
 Wolibner, W. 204, 275 [171]
 Wu, J. 204, 270 [40]; 270 [41]
 Wu, S. 268, 275 [172]
 Wzrosek, D. 176, 193 [119]

 Xin, Z. 206, 275 [173]

 Yamamoto, M. 574, 583, 588, 589, 595, 618 [98]
 Yang, T. 472, 473 [19]
 Yao, P.-F. 548, 620 [162]
 Yao, Z.-A. 472, 473 [19]
 Yau, H.-T. 260, 275 [174]
 Yi, T.M. 167, 168, 193 [120]
 Yong, J. 514, 526 [53]
 Young, R.M. 550, 552, 620 [163]
 Yudovich, V.I. 204, 267, 275 [175]

 Zaag, H. 174, 189 [24]
 Zabczyk, J. 113, 134 [35]
 Zeytounian, R.K. 201, 275 [176]; 275 [177]
 Zhang, L. 206, 275 [173]
 Zhang, X. 530, 548, 556, 561, 579, 583Ð585, 598, 600Ð606, 612, 613, 616 [53]; 618 [121]; 619 [146]; 620 [156]; 620 [164]; 620 [165]; 620 [166]; 620 [167]; 620 [168]; 620 [169]
 Zheng, Y. 402, 475 [82]; 475 [83]
 Zhidkov, P. 402, 469, 476 [114]
 Zigmund, S.H. 146, 157, 193 [121]
 Zuazua, E. 530, 531, 547, 548, 551, 552, 554, 559Ð562, 564, 569, 570, 572, 573, 575Ð579, 581Ð585, 587, 588, 593, 597, 598, 600Ð607, 609, 611Ð613, 615 [22]; 615 [25]; 615 [26]; 615 [27]; 615 [30]; 616 [40]; 616 [41]; 616 [44]; 616 [48]; 616 [50]; 616 [53]; 616 [56]; 616 [57]; 616 [58]; 616 [62]; 617 [69]; 617 [70]; 617 [71]; 617 [86]; 617 [89]; 618 [103]; 618 [109]; 618 [110]; 618 [119]; 618 [121]; 618 [122]; 619 [123]; 619 [125]; 619 [126]; 619 [127]; 619 [128]; 619 [146]; 620 [161]; 620 [165]; 620 [166]; 620 [167]; 620 [168]; 620 [169]; 620 [170]; 620 [171]; 620 [172]; 620 [173]; 620 [174]; 620 [175]; 621 [176]; 621 [177]; 621 [178]; 621 [179]; 621 [180]; 621 [181]
 Zuily, C. 556, 620 [148]
 Zwanzig, R. 148, 193 [122]
 Zworski, M. 548, 615 [21]

Subject Index

- $\mu \sqsubset A$ 285
- $\partial^* A$ 289
- ε -neighborhood 542
- A
 - absolutely continuous part 283, 286
 - activated pathways 158
 - activator–inhibitor systems 161
 - adaptation 140, 142, 161, 167
 - perfect 164
 - adjoint
 - equation 535
 - system 532, 533, 537, 543, 545, 553, 555, 568
 - admissible controls 499, 537
 - Airy equations 552
 - Alberti’s
 - lemma 317
 - rank-one theorem 290
 - algebraic Kalman condition 537
 - Ambrosio’s renormalization theorem 281, 308
 - amplification 142, 159
 - analytic coefficients 542, 556
 - Anzellotti’s weak trace 350
 - approximate
 - controllability 532, 542, 562, 563
 - discontinuity set 287
 - unit normal 289
 - area formula 12
 - asymptotic
 - behavior 529, 530
 - gap 550
 - problems 197, 198, 216, 245, 251, 268
 - attractors 562
- B
 - $B^+(x)$ 287
 - $B^-(x)$ 287
 - $\tilde{B}(x)$ 287
- backward
 - scheme 522
 - uniqueness 563
- bang-bang controls 568
- bicharacteristic ray 547
- bilinear control 529
- blow up 392, 393, 402, 403, 407, 417, 434, 435
- Boltzmann equation 386
- Bouchut’s lemma 317
- boundary
 - control 542, 557
 - controllability 546
 - estimate 546
 - observability 547
 - stabilization 560
 - traces 542
- Bressan’s compactness conjecture 280, 364
- BSDE 480, 514
- 2BSDE 515
- BV
 - functions 288
 - structure theorem 288
- C
 - Caccioppoli sets 289
 - cAMP 140, 165, 166, 171, 181, 188
 - Cantor part 283, 286
 - Carleman
 - approach 548
 - inequalities 548, 569
 - Casimir functional 440, 441, 454, 455, 469, 472
 - Cayley–Hamilton theorem 535
 - cell
 - membrane 141, 142, 166
 - motility 141
 - navigation 153
 - chain rule 57, 67
 - change of variables for traces 357
 - chemotactic
 - cells 140, 141
 - collapse 173, 174
 - units 141
 - chemotaxis
 - and biased random walks 175
 - and stochastic many-particle system 178
 - and velocity jump processes 177

- closed graph theorem 535
- closed-loop 559
- coarea formula 289
- coercivity 534, 538
- commutator 309
 - estimate 312, 369
- comparison for viscosity solutions 519
- compensator 491, 509
- compressible fluid 197, 217
- continuation criterion 392, 397, 401, 413, 416, 423
- continuity equation 19, 29, 85, 108
- continuous observability 543
- control
 - map 536
 - of minimal norm 536
 - region 546
 - theory 529
 - to trajectories 577
- controllability
 - problem 529
 - property 538
- controllable 532, 533
 - pairs 537
- controller 536
- controls of minimal norm 563
- convergence
 - in the Wasserstein space 15
 - narrow 10
- convexity along geodesics 39
- Coron's return method 529
- cost of approximate controllability 542, 564
- crosstalk 160
- D
 - $D^a B$ 288
 - $D^c B$ 286, 288
 - $D^j B$ 286, 288
 - $D^s B$ 286
- D'Alembert formula 557, 608
- damper 561
- density of the range of the semigroup 564
- detection
 - spatial gradients 149, 152
 - temporal gradients 151, 152
- Dictyostelium discoideum* (*Dd*) 140, 141, 165, 168, 170
- differentiability
 - of gradient flows 90
 - of W_2 37, 81
- differential games 496
- diffusion 142, 149
 - effects on ligand binding 147
 - equations 129
- direct method of the calculus of variations 534, 538
- Dirichlet
 - boundary condition 492
 - Laplacian 549
 - problem 563
- discrete Markov processes 488
- displacement convexity 39
- distributed parameter systems 529
- divergence problem 365
- domain of the operator 561
- double averages lemma 370
- duality 563
- dust 402, 403
- dynamic programming
 - equation 501
 - principle 497
 - abstract 500
- dynamical properties 530
- E
 - effective medium 148
 - eigenfunction estimate 549
 - eigenfunctions 549
 - elliptic operator 573
 - energy 547
 - dissipation law 602
 - functional 441, 455, 472
 - interaction 42
 - subdifferential of 77
 - internal 42
 - subdifferential of 70
 - potential 41
 - subdifferential of 69
 - space 536, 553
 - energy-Casimir functional 440–442, 462, 465, 472
 - entropy 47, 113
 - solution 323
 - subdifferential of 75
 - entropy–entropy flux pair 321, 323
 - equilibrium 542
 - equipartition of energy 560
 - escape functions 547
 - Escherichia coli* (*E-coli*) 140, 160
 - essential boundary 289
 - Euler equations 530
 - Euler–Lagrange
 - equation 446, 448, 460, 470, 555
 - system 402, 403, 405, 406, 414, 442, 443, 460, 470–472

- exact controllability 532, 542, 543, 562
- excitability 188
- exit time 493
- explicit bounds 548
- exponential decay 560, 561
 - rate 602
- exterior boundary 547
- F
 - $f_{\#}\mu$ 285
- feedback 559
 - control 167
- Feynman–Kac formula 479, 481
- final
 - condition 532, 534
 - datum 532
 - target 536, 553
- finite
 - speed of propagation 545
 - velocity of propagation 537
- finite-approximate controllability 554, 555, 563
- finite-dimensional systems 529, 542
- finite-energy solutions 542
- Fisher information 113
- Fisk–Stratonovich integral 515
- fixed point 529, 548
- fluctuations 142
 - kinetic binding 145
 - of ligand 144
- fluid mechanics 201, 206, 245, 261
- fluid–structure interaction 530
- fluids 530
- Fourier
 - series 563
 - space 547
- fractional parabolic equation 573
- Fréchet subdifferential 55, 58
 - closure of 57, 64
 - minimal selection 57, 66
 - monotonicity 56, 62
 - of convex functions 56, 62
 - perturbations of 56, 61
- free boundary problems 530
- Fubini’s theorem for traces 354
- functional setting 542
- functions of bounded variations 288
 - special 288
- G
 - gap condition 549, 550
- Gaussian
 - beam 546, 561, 599
 - heat kernel 570
- generating function 145
- geodesics 16
- geometric
 - control condition 546
 - optics 546
 - PDE 511
 - restrictions 546
- Ginzburg–Landau equation 182
- global
 - existence 392, 393, 402, 403, 405, 407, 413, 414, 416, 417, 430, 431, 433, 469
 - gap 551
- globally Lipschitz nonlinearities 564
- gradient
 - chemical 141
 - flows 84
 - existence of 95
 - properties of 89
 - uniqueness of 88
- H
 - Hahn–Banach theorem 554, 564
- Hamiltonian 440
 - system 547
- heat equation 529, 530, 536, 562, 563
- Hilbert uniqueness method 543
- Holmgren’s uniqueness theorem 554, 556, 563
- homogenization 530
- Hopf bifurcation 183, 184
- hydrodynamics 197, 198, 268
- I
 - infinite-dimensional systems 536
 - infinitesimal generator 482
- Ingham inequality 550
- initial
 - data 529, 536
 - value problem 487
- integro-differential equation 490
- interface 530
- internal
 - variables 178
 - controllability 542
- inviscid limit 197, 198, 201, 202, 205, 212, 244, 268
- irreversible models 529
- J
 - J_{μ} 286
 - J_B 287
- jump
 - part 283, 286
 - set 283, 287

K

- Kalman
 - condition 532, 535
 - rank condition 533, 537
- Kannai transform 572
- Kantorovich problem 14, 17
- Keller–Segel model 168, 170, 174–177, 180
- Keyfitz and Kranzer system 279, 321
- kinases
 - activated 157
 - proteins 157
 - signal
 - amplitude 158
 - duration 158
 - signaling time 158
- kinetic equation 145, 386, 387, 392, 408
- Kolmogorov–Fokker–Planck equation 113
- Kruzhkov solution 322

L

- λ – ω systems 182
- lateral boundary 542
 - condition 503
- Lebesgue limit 287
 - one-sided 287
- level set equation 511
- lifting argument 545
- ligand
 - binding 141, 142
 - concentration 143
 - multivalent 156
- ligand–receptor aggregates 155
- linear
 - problems 529
 - semigroup 481
 - stability 171
- linearization 548
- Lipschitz continuous 548
- local
 - existence 392, 393, 402, 416, 430
 - null controllability 530
- locality property (of weak traces) 352
- locally distributed damping 562
- log-concavity 50
- logarithmic
 - decay 561
 - gradient 114
- lower semicontinuity 566
- Lyapunov functional 547, 561

M

- macroscopic behavior 168
- mass constraint 441, 451, 464

- mass-Casimir constraint 441, 455, 464
- master equation 145, 175
- maximum principle 484
- mean curvature flow 511
- method of
 - multipliers 544, 545
 - transposition 543
- metric derivative 13
- microlocal
 - analysis 546, 549
 - approach 546
 - propagation 547
- microscopic behavior 168
- minimization 555
- minimizer 534, 536, 537
- minimizing movements 95
- moment problem 552, 569
- Monge problem 14
- monotone semigroup 483
- Monte Carlo method 521
- multiplier techniques 546
- multistructures 559

N

- N -body problem 385, 387, 392, 414
- narrow convergence 10
- Nash–Moser’s iteration 529
- Navier–Stokes equations 198, 199, 201, 202, 204–206, 212, 215–217, 221, 223, 227, 229, 233, 243, 249, 530
- nearly incompressible fields 294
- networks 551
- Neumann condition 493
- neural navigation 140
- nodal sets 609
- nonharmonic Fourier series 549, 550
- nonhomogeneous boundary value problems 542
- nonlinear
 - damping 559
 - stability 386, 439, 466
 - systems 529
- normal derivative 546
- null controllability 542, 562, 563, 568
- numerical approximation 530

O

- observability 532
 - constant 536, 543, 548
 - inequality 534–536, 538, 543, 546, 549, 555, 568
- observable 533
- observed quantity 543
- obstacle problem 510

one space dimension 569
 open-loop 559
 optimal
 control 496
 of diffusion 501
 problems 530
 transport maps 17
 ordinary differential equations 529
 orthogonal projection 565
 orthonormal basis 549
 oscillatory media 182
 overdamping 599

P

parabolic
 models 530
 PDE 484
 parabolic problems 529
 partial differential equations 529
 partial measurements 536
 pattern
 formation 141
 in Dd 180
 stabilization 163
 periodic rays 547
 phosphorylation 142, 157
 plans 9
 plant 554
 plate equations 529
 propagation of singularities 546
 push-forward of a measure 285

Q

quadratic functional 533, 563
 quasibang-bang controls 556, 608
 quasivariational inequality 510

R

random walk 175
 range of the semigroup 563
 rate of decay 602
 Rauch's commutator condition 335
 ray 546
 reachability set 511
 reachable 563
 data 542
 states 532
 reaction-diffusion equations 183
 real exponentials 569
 receptor 142
 clustering 143, 154, 157
 by bivalent ligands 154
 occupancy 142, 160

rectifiable sets 288
 regular
 functionals 65
 Lagrangian flow 291
 relaxed invariance principle 602
 renormalization property 296, 297
 renormalized
 entropy solution 280, 322
 solutions 281
 Riccati operator 560
 rigid body 530
 robustness 538, 554
 rotating fluid 237, 238, 244
 running cost 499

S

S_B 287
 SBV 288
 Schrödinger equation 529, 550, 552
 semigroup 560, 563
 semilinear
 models 529
 wave equation 560
 sensitivity 139, 140, 142, 157
 sharp observability inequalities 550
 signaling
 networks stability 160
 pathways 157
 similarity variables 573
 simply coupled diffusion 489
 singular
 control 510
 limit 198, 201, 217, 237, 268
 Skorokhod problem 493
 slope 55, 60
 lower semicontinuity of 57, 63
 of convex functionals 56, 62
 of regular functionals 57, 66
 spectral
 abscissa 562
 characterization 549
 estimates 548
 properties 529
 spectrum 562
 spirals 181, 183, 188
 stabilization 546, 559
 state
 equation 536, 542
 space 553
 stationary
 problem 494
 solution 386, 433, 436

- steady state 406, 431, 435–442, 460, 462, 465, 468–471
- Strichartz estimates 561
- strong trace 322
- structures 530
- supercritical nonlinearities 561
- switching 538
 - control 510
- system of thermoelasticity 569
- T
- tangent space 34, 38
 - and strong subdifferentiability 58
- tangential set of a BV vector field 350
- target 181, 542, 555
 - problem 512
- tightness 10
- time-irreversibility 568
 - equations 536
- trace 480
- transition probability 488
- transport of measures 9
- trigonometric polynomials 552
- U
- uncontrolled dynamics 536
- uniform exponential decay 559
- uniformly integrability 11
- unique continuation 529, 538, 553, 555
 - property 543
- uniqueness property 535, 536
- V
- value function 496, 500
- variable coefficients 547, 548
- variation of constants formula 532
- variational
 - approach 554
- integrals 68
- methods 529
- problem 440–443, 449, 454, 455, 460, 462, 464, 470
- vasculogenesis 176
- velocity
 - feedback 560
 - of propagation 537
- verification theorem 508
- viscosity solution 501
- Vlasov equation 386, 391, 394, 413, 414, 423, 424, 430, 431, 436, 438, 440, 462
- Vlasov–Einstein system 386, 392, 401, 413, 430, 469
- Vlasov–Fokker–Planck equation 430
- Vlasov–Fokker–Planck–Poisson system 430
- Vlasov–Maxwell system 386, 392, 401, 407, 413, 430
- Vlasov–Maxwell–Boltzmann system 386
- Vlasov–Nordström system 430
- Vlasov–Poisson system 386, 387, 390, 392, 393, 400, 402–405, 407, 408, 416, 417, 429–433, 435–440, 442, 460, 462, 466, 468–472
 - relativistic 391, 392, 407, 417, 433, 435
- Vlasov–Poisson–Boltzmann system 386
- Vol’pert chain rule 289, 290
- W
- Wasserstein distance 14
 - differentiability of 81
 - semiconcavity of 46
 - slope of 81
- wave
 - equation 529, 530, 537, 542, 549
 - packet 549
- wave-like equations 546
- waves on networks 552
- weak continuity of traces 354